Edith Cowan University

Copyright Warning

Digital representation for assessment of spoken EFL at university level: A case study in Vietnam

This thesis is presented for the degree of

Doctor of Philosophy

Thi Bich Hiep Vu

Edith Cowan University

School of Education

2021

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

ABSTRACT

Assessing the speaking performance of students who are studying English as a Foreign Language (EFL) has mainly been conducted with face-to-face speaking tests. While such tests are undoubtedly interactive and authentic, they have been criticised for subjective scoring, as well as lacking an effective test delivery method and recordings for later review.

Technology has increasingly been integrated into speaking tests over the last decade and become known as computer-assisted or computer-based assessment of speaking. Although this method is widely acknowledged to measure certain aspects of language speaking effectively, such as pronunciation and grammar, it has not yet proved to be a successful option for assessing interactive skills. An effective testing method is deemed to maintain the interactivity and authenticity of live speaking tests, able to deliver tests quickly and efficiently, and provide recordings of performances for multiple marking and review.

This study investigated digital representation of EFL speaking performance as a viable form of student assessment. The feasibility of digital representation has previously been examined in relation to authenticity and reliability in assessment of different subjects in Western Australia, including Italian, Applied Information Technology, Engineering Studies, and Physical Education Studies. However, as far as the researcher is aware, no studies have yet assessed EFL speaking performance using digital representation. In an attempt to bridge this gap, this study explored the feasibility of digital representation for assessing EFL speaking performance in a university in Vietnam, the researcher's home country.

Data collection was undertaken in two phases using a mixed methods approach. In Phase 1, data related to English teachers' and students' perceptions of Computer-Assisted English Speaking Assessment (CAESA) were collected. Their perceptions were analysed in relation to the outcomes of a digital speaking assessment trial using the Oral Video Assessment Application (DMOVA). In Phase 2, student participants took an English speaking test while being videoed and audio recorded. English teachers invigilated and marked the trial test using the current method, followed by the digital method. Data were collected via Qualtrics surveys, interviews, observations and databases of student performance results. The feasibility of digital representation in

assessing EFL speaking performance was analysed according to the Feasibility Analysis Framework developed by Kimbell, Wheeler, Miller, and Pollitt (2007).

The findings from Phase 1 indicated that both teachers and students had positive attitudes towards computer-assisted assessment (CAA). They were confident with computer-assisted English assessment (CAEA) and preferred this testing method to the current paper-and-pencil process. Both cohorts believed that CAEA enhanced the precision and fairness of assessments and was efficient in terms of resources. However, some participants were sceptical about the authenticity of computer-assisted EFL speaking tests because it failed to foster conversations and interactions in the same way as face-to-face assessments. In spite of their scepticism, teachers and students indicated their willingness to trial DMOVA.

Phase 2 identified the feasibility dimensions of DMOVA. This method of digital assessment was perceived to enhance fairness, reliability and validity, with some correlations between the live interview and digital tests. Teachers found it easy to manage the speaking tests with DMOVA and recognised the logistical advantages it offered. DMOVA was also credited with generating positive washback effects on learning, teaching and assessment of spoken English. In addition, the digital technology was compatible with the existing facilities at the university and required no support or advanced ICT knowledge. Overall, the benefits of the new testing method were perceived to outweigh the limitations.

The study confirmed that digital representation of EFL speaking performances for assessment would be beneficial for Vietnam for the following reasons: (a) it has potential to enhance the reliability and accuracy of the current English speaking assessment method, (b) it retains evidence of students' performance for later assessment and review, and (c) it facilitates marking and administration. These changes could boost EFL teaching, learning, and assessment, as witnessed in the trial, leading to increased motivation of teachers and students, and ultimately, enhancement of students' English communication skills. The findings of the study also have implications for English speaking assessment policies and practices in Vietnam and other similar contexts where English is taught, spoken and assessed as a foreign language.

DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

i. Incorporate without acknowledgment any material previously submitted for a degree or diploma in any institution of higher education,

ii. Contain any material previously published or written by another person except where due reference is made in the text of this thesis, or

iii. Contain any defamatory material

Signature  Date 10 April 2021

ACKNOWLEDGEMENTS

I would like to express my most sincere gratitude to my supervisors, Dr Anne Thwaite, Dr Jeremy Pagram, and Dr Alistair Campbell, who always gave me enlightening guidance, kindest support and extensive encouragement during all the ups and downs of my doctoral journey. My supervisors inspired and lifted me up and helped me grow academically and intellectually. I am very happy, lucky and proud to have studied under their supervision.

I would like to thank Dr Henny Nastiti for sharing her expertise and giving me tremendous mentoring and unconditional help. She was like my big sister who was always close to me, willing and ready to answer all of my questions, and gave me good advice to help me solve my problems. I wish to thank Dr Jo McFarlane and Ms Bev Lurie for their time and their kind help to proofread this thesis. They worked closely with me to clarify my ideas and guide me on how to give them life in terms of writing style and expression.

I would especially like to thank the staff members at GRS, Edith Cowan University, and I would also like to thank staff members in the library at Mt Lawley campus. All of you have been there to support me in my search for literature for my PhD thesis.

I would like to acknowledge the financial support provided by the VIET-Joint Scholarship which offered me a great opportunity for my higher study and made my dream come true.

I would like to thank my friends Dr Thi Thu Lan Nguyen, Dr Phan Thu Ngan Nguyen, Ms Thi Hien Tran, Ms Zina Cordery, and Dr Huifen Jin for their kind support, encouragement and friendship, which created a source of positive energy for me to recover from all my hardships and look ahead to the success of today.

Especially, I would like to show my special sincere thanks to my husband for his understanding and caring, which brought me happiness and motivation to complete the biggest learning course of my life. This thesis would not have been completed without the encouragement and motivation I got from my kids, who were so caring and loving, and from my sister and brothers, who always gave me encouragement and support. I especially would like to thank my Mum, a retired secondary teacher, who closely observed every one of my steps and gave me unconditional love and support. I also know that my dear passed-away Daddy always follows and supports me even when he

is not in this world. I was motivated so much in my study and learnt how to turn loss into gain and turn misfortune into my success today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS, ABBREVIATIONS AND DEFINITIONS

Acronyms and abbreviations

CAA	Computer-Assisted Assessment
CAEA	Computer-Assisted English Assessment
CAESA	Computer-Assisted English Speaking Assessment
CALA	Computer-Assisted Language Assessment
CALL	Computer-Assisted Language Learning
CASA	Computer-Assisted Speaking Assessment
CBA	Computer-Based Assessment
CEFR	The Common European Framework of Reference for Languages
CLT	Communicative Language Teaching
CMS	Content Management System - a university intranet
COPI	Computerised Oral Proficiency Instrument
CSA	Computer-Supported Assessment
CSaLT	Centre for Schooling and Learning Technologies
EF EPI	Education First English Proficiency Index
EFL	English as a Foreign Language
ELF	English as a Lingua Franca
ELSA	English Language Speech Assistant
ELT	English Language Teaching
ESP	English for Specific Purposes
FPT University	Financing and Promoting Technology University
ICT	Information and Communication Technology
IELTS	International English Language Testing Systems
LAD	Language Acquisition Device
MALA	Mobile-Assisted-Language Assessment
MOET	(Vietnamese) Ministry of Education and Training
NFLP/ 2020 Project	National Foreign Languages Project 2020
NLP	Natural Language Processing
OPI	Oral Proficiency Interview
OVA App	Oral Video Assessment Application

PDA	Personal Digital Assistant
SLA	Second Language Acquisition
SOPI	Simulated Oral Proficiency Interview
SPSS	Statistical Package for the Social Sciences
S-R-R	Stimulus, Response, and Reinforcement
TAM	Technology Acceptance Model
TOEFL	Test of English as a Foreign Language
TOEFL iBT	TOEFL internet-Based Test
TOEIC	Test of English for International Communication
VOCI	Video Oral Communication Instrument

Definitions

1400/QD/TT	The Decision 1400 by the Prime Minister of the Vietnamese government issued on 30 September 2008 named “Teaching and Learning Foreign Languages in the National Education System, Period 2008-2020”.
Curriculum	Referring to the lessons and academic content taught in a school or in a specific course or program.
DMOVA	Digital speaking assessment method using Oral Video Assessment Application.
Digital representation of student performance	Electronic files of student performances recorded in forms of audio, films, text and/or graphics, and photographs.
Functional dimension	Regarding the validity and reliability of digital representations for assessment and their comparability with other assessment methods.
Manageability	The practicalities of administration, collection and assessment of student work in digital forms.
NVivo	A qualitative data analysis computer software package produced by QSR International.
Pearson PTE Academic tests	Computer-based exams.
Pedagogy	The method or practice of teaching.
Pedagogy of digital form of assessment	The extent to which digital representations for assessment can support and enhance teaching and learning.
Technology dimension	The extent to which existing technologies are suitable for adaptation to the purposes of assessment.

Washback effect	Referring to the impact or influence of assessment practices on all individuals involved in the teaching-learning process.

CHAPTER 1
INTRODUCTION

Overview

This study presents the results of a four-year research project exploring the feasibility of using digital representation for English as a foreign language (EFL) speaking assessment in a university context in Vietnam. The digital representation involved the process of recording students' performances to allow multi-marking and facilitate reviewing the results. This new digital testing method also modified the way language teachers marked students' English speaking skills. Instead of giving a live judgment in real time, dependent on the teacher's memory and the potential influence of student impressions, teachers were able to review student performances at their convenience and compare and contrast with the results of others before determining the final outcome.

Since the advent of computers, their integration in teaching and assessment has been extensively and intensively researched for the purpose of enhancing effectiveness and reliability. However, there is one aspect of English language teaching (ELT) that has not changed greatly over time – the assessment of students' speaking performance. Oral proficiency or spoken language seems to be the most difficult aspect of the language repertoire to assess. For a long time, face-to-face interviews have been viewed as the best way to demonstrate communicative skills and fully assess the richness of communicative competence. However, this may be outdated, given that computers have been well integrated into speaking assessment and proven to provide higher levels of practicality and reliability.

Conventional face-to-face interviews undeniably possess distinct constructs for assessing spoken language (Bernstein, Moere, & Cheng, 2010). However, interviews have limitations in terms of reliability, validity, impact and feasibility (Margaret & Megan, 2010). In regard to reliability, testers inevitably make mistakes from time to time, thereby posing threats to consistency. Double-rated oral proficiency interviews have been credited with higher reliability, but local and unofficial single-rated interviews may be less reliable (T. Cox & Davies, 2012; Margaret & Megan, 2010). The time is ripe for a new digital performance testing approach that takes advantage of the functionality offered by computers and the internet, suited to a new generation of

students. It is also time for universal assessment of speaking performance to supplant locally accepted methods (Margaret & Megan, 2010; Moere, 2010).

Currently, speaking tests are low-tech, costly, time-consuming, subjective and unreliable. Testing and marking can only be undertaken by teachers or specialists in the target subject, creating difficulties when qualified teachers are unavailable. Integrating ICT into speaking tests can help improve the quality of testing by eliminating problems associated with conventional assessment methods.

Researchers have been persistent in their quest for a more effective and reliable method of speaking assessment. McNamara (2000) suggested a "semi-direct test" (p. 83) that allows test-takers to respond to questions while their performance is tape-recorded and assessors mark from the tape. This testing method is believed to be fairer and more economical with a large number of test-takers, because it reduces the administrative work and requires less involvement by interlocutors or interviewers. Although test-takers respond to the same questions, they experience different feelings about the recordings. Some feel comfortable speaking in front of a machine, while others feel constrained and voiceless. The tests are often not as economical as once believed, due to expensive equipment and time-consuming preparation. McNamara (2000) claimed: "In the dazzle of technological advance, we may need a continuing reminder of the nature of communication as a shared human activity, and that the idea that one of the participants can be replaced by a machine is really a technological fantasy" (p. 85).

Feasibility of the Computerised Oral Proficiency Instrument (COPI) was also investigated by Larson (2000), who found a number of benefits. First, the quality of sound generated by computers was better than the old technologies, like audio cassette tapes. Second, the method offered extreme flexibility for retrieving recorded oral performances, allowed markers to focus on the essential elements to be assessed, ignored warm-up responses, and reduced marking time. COPI programs also contain different forms of instructions, such as audio, video clips, cartoons, and charts, all of which are simple and comprehensible.

WhatsApp, a social networking application on smartphones, and an e-portfolio have also been investigated for assessing students' English speaking competence (Tarighat & Khodabakhsh, 2016). Described as Mobile-Assisted-Language Assessment (MALA), this method allowed students to study while they were being assessed and enabled peer-checking amongst test takers. All participants' speaking performances were recorded

and posted on the social networking platform; participants viewed the recordings on their smartphones and added comments to their friends' speaking performances. Teachers made the final comments, resolved all disagreements about specific aspects of the recordings, and provided a final score. Although MALA created opportunities for peer-checking, self-checking and fairer assessment of students' oral performances, wayward students could cheat and some students received negative comments from others. Nevertheless, MALA was recommended for homework tasks and as an additional tool for official assessments (Tarighat & Khodabakhsh, 2016).

Another study on assessing learners' practical performance was conducted in Western Australia by Newhouse and Cooper (2013). It was a part of a three-year study that used digital assessment to evaluate Italian oral performance in summative tests. It included different approaches, such as "a portfolio of sub-tasks leading up to a video-recorded oral presentation, a computer-based exam, a video recorded interview, and an online exam that included oral audio-recordings" (p. 321). The study indicated a preference for using digital methods to assess oral performance rather than conventional face-to-face methods. Marking by means of the digital method was thought to be equally reliable and valid as the conventional method, as well as faster and more convenient. However, some technical complexities, unfamiliarity with the digital testing method, and nervousness and anxiety in front of the camera appeared to dampen teachers' and students' enthusiasm for the digital method. Newhouse and Cooper (2013) recognised the potential of this new method and stated that computer-based oral tests are manageable and feasible. They recommended further study in different contexts.

Digital representation seems to be a promising method of assessing performance. In the e-scape project in the United Kingdom, Kimbell et al. (2007) studied the use of digital cameras to record and display students' performance on a web space accessible to students, teachers and assessors. Stables and Kimbell (2007) claimed that the digital representation of students' performance provided evidence of assessment and engaged and motivated students. Their study showed that digital representation provided a repository of students' work and awoke student reflection and critical input from teachers.

A reliable method of speaking assessment with digital technologies is long overdue to bring speaking skills onto an equal footing with reading, writing and listening in school tests and examinations. Teachers and students may be more encouraged to teach and learn speaking skills, with the overall aim of improving the English communication

skills of 21st century students (Greenstein, 2012) in particular and English learners in general.

The current study addressed this goal at FPT University in Vietnam, by combining digital technologies with English speaking assessment to measure validity and reliability in the latter. It examined correlations between live and digital marking and identified strengths and weaknesses in the new testing method, from which flowed recommendations for further study.

This introduction includes an overview of EFL education in Vietnam and discusses EFL teaching and learning at tertiary level, as well as the challenges of EFL assessment. The chapter also presents the particular context of the study, the purpose, significance, scope, research questions and organisation of the thesis.

Background

English Language Education in Vietnam

The increasing role of English as a means of international communication has promoted the teaching and learning of English in non-English speaking countries to boost their socio-economic development and globalisation. In this climate of internationalisation for economic development and cultural exchange, the demand for high-level English communication skills among younger generations is higher than ever. Vietnam is an active participant in this trend to enhance the teaching and learning of English. Although the position and status of English in the Vietnamese school curriculum has changed throughout history, English is currently the most important foreign language at all school levels and a compulsory subject in the education system (Hoa & Tuan, 2007).

Little is known about the introduction and earliest teaching of English in Vietnam, because no written documents or official English textbooks have ever been found. During wartime, prior to 1975, the status of English differed in schools in the north and south of Vietnam. Before 1986, teaching and learning English was limited to some schools due to the dominance of Russian (Hoang, 2010). Since economic reform in 1986, English has become the foremost foreign language taught in Vietnam (Hoang, 2010; Ngan, 2012) and is believed to provide significant opportunities for employment, promotion and further education. English proficiency is fast becoming a prerequisite for job recruitment and entry into higher education. Learners do not merely learn English for employment opportunities, but also for personal enrichment (Shukla, 2018). It is understood that the English competence of Vietnamese citizens contributes significantly

to national socio-economic development and international integration, and therefore, English education receives more attention in the educational policies of the Vietnamese government than ever before.

The Education First English Proficiency Index (EF EPI) is a ranking system of countries based on the average level of English skills of adult learners taking English tests online. EF EPI is the product of Education First, an international education company established in 1965. To be included in the index, countries must have at least 400 test takers. Scores are calculated based on the results of the EF Standard English Test (EF SET) for a maximum of 100 points. According to the 2018 EF EPI (EPI, 2018) results, Vietnam ranked 41 among 88 countries and territories worldwide, classified as moderate level. Vietnam was placed 14th out of the 17 countries listed at the moderate level, equivalent to level B1 of the Common European Framework of Reference for Languages (CEFR). In Asia, Vietnam ranked 7 out of 21 with a score of 53.12, behind the Philippines and Malaysia in the same region, while the average score for Asia was 53.49.

Table 1.1

EF English Proficiency Index

Year	**EF EPI Ranking**	**EF EPI Proficiency Bands**	**Asia EF EPI Ranking**	**EF EPI Score**
2014	33/63	Moderate	9/14	51.57
2015	29/70	Moderate	9/16	53.81
2016	31/72	Moderate	7/19	54.06
2017	34/80	Moderate	7/20	53.43
2018	41/88	Moderate	7/21	53.12

The above numbers show that the English proficiency levels of the Vietnamese people increased in 2018 (EPI, 2018) compared to 2014 (EPI, 2014). However, the country's ranking dropped in 2018 compared to 2016 (EPI, 2016), with a score of 54.06. Overall, the EF English Proficiency Index for Vietnam over the five-year period, from 2014 to 2018, shows little improvement, despite the government's 450 million USD investment in language learning between 2008 and 2020, with 85% of the budget allocated to teacher training (EPI, 2014, p. 15). However, the actual results achieved from this huge investment in English teaching and learning have been less positive than expected: "Many school leavers cannot read simple texts in English nor communicate with English speaking people in some most common cases" (Le, 2013, p. 66).

Previous studies showed that many factors affected the quality of English teaching and learning in Vietnam. These were identified as large class sizes, insufficient time and authentic contexts for communicative practices, teaching for examinations, teachers' limitations in the use of technologies to aid teaching, and poor teaching resources (Hoang, 2008; Le, 2013; H. T. Nguyen, Warren, & Fehring, 2014; V. L. Nguyen, 2010; Tran, 2013). Moreover, Le (2013) pinpointed language testing and assessment as important factors affecting the quality of EFL teaching and learning in Vietnam and claimed that they were not effectively facilitating the learning and teaching of English language skills. Assessment was blamed for an imbalance in teaching and learning English communication skills, due to the lack of speaking and listening tests and examinations. A mismatch between language teaching and testing was also cited as a barrier to EFL learning and teaching in Vietnam (Hoang, 2010), since English was taught by means of Communicative Language Teaching, yet English tests focused on vocabulary and grammar (Hoang, 2010; Le, 2013; Tran, 2013).

The Vietnamese government issued numerous policies designed to enhance the quality of English teaching and learning across the entire education system. In particular, the Decision 1400 (1400/QD/TT) was issued by the Prime Minister on 30 September 2008 and named "Teaching and Learning Foreign Languages in the National Education System, Period 2008-2020". The Decision stated that, by the year 2020, most young Vietnamese graduates should be able to use a foreign language independently and confidently in communication. It also focused on solutions to address persisting issues in English testing and assessment.

Teaching and learning EFL received even more attention after the proclamation of the National Foreign Languages Project 2020 (NFLP/ 2020 Project) by the Ministry of Education and Training. The aim of the 2020 project was for most Vietnamese students to be able to confidently use a foreign language, primarily English, in their daily communication, study and work by 2020. To achieve these goals, MOET focused on "improving quality of education through renovation of curriculum, textbooks, teaching methods, teacher training and development" (Huong, 2010, p. 111). However, the mismatch between English teaching and testing still needed to be resolved (Hoang, 2010) and required "macro-changes including reforming the current grammar-based testing system" (V. T. Nguyen & Ngo, 2015, p. 1840).

In summary, English is the most important foreign language taught and learnt in the education system in Vietnam today, because it has become "an indispensable language

for intra-national communication and international communication" (Ngan, 2012, p. 265). The Vietnamese government prioritised EFL teaching and learning by issuing favourable policies and investing extensively. However, on a macro level, the quality of EFL teaching and learning in Vietnam still needs further improvement, since English proficiency is limited, and solutions are needed to address the hindrances.

English Tertiary Education in Vietnam

Hoang (2010) described tertiary English language teaching in Vietnam in two ways. The first is where English is taught as a discipline for students who aspire to becoming English teachers, translators or linguists; these students learn English as a major subject at university. The second is where English is taught as a normal subject at university to all non-English major students. This study focused on the second type – English for non-major English students.

Underpinned by the belief that "tertiary education is a key indicator of a nation's effort to develop a highly skilled workforce needed to compete in today's global economy" (Linh, Thuy, & Long, 2010, p. 4), English is fundamental for internationalising higher education in Vietnam (Duong & Chua, 2016). Together with the early introduction of English in primary schools, English education at tertiary level also received priority from the Vietnamese government, through ambitious investment to transform English teaching and learning (H. T. Nguyen, Fehring, & Warren, 2014). Together with others, the National Foreign Languages Project 2020 (NFLP/2020 Project) was targeted to improve students' English proficiency, while the Government 911 Project focused on training tertiary teachers – these initiatives are just some examples of the Vietnamese government's efforts to enhance the quality of teaching and learning at tertiary level.

Different approaches and technologies have been applied over the years to improve language teaching and enhance learners' competence (V. L. Nguyen, 2010; Thao & Le, 2011). For example, the Communicative Language Teaching method was adopted to provide a student-centred, rather than teacher-centred approach (H. T. Nguyen, Fehring, et al., 2014). Nevertheless, the quality of EFL teaching and learning at Vietnamese universities still fail to meet expectations (Tran, 2013) and remain a challenge in tertiary education. Despite its importance to students' future study and work, English has been poorly taught at universities and the outcomes lower than expected (Tran, 2013), as evidenced by the elementary levels of English communication skills (Hoang, 2008) among Vietnamese graduates. Hoang conducted an English proficiency test that was

randomly extracted from the Key English Test (KET), one of the Cambridge English exams, and found 20% of student participants scored below 5/10. Thirty percent of students passed the English speaking and listening tests, and only one student achieved 7.5/10 for speaking skills. One of the factors found to hinder students' communication skills was the absence of English speaking tests at non-English major universities in Vietnam; most universities designed English achievement tests to check students' grammar and sentence structure without checking their writing, speaking and listening skills (Hoang).

The lack of a speaking component in EFL tests and examinations has also affected the efficacy of English learning and teaching. "Of the challenges that teachers face, the exam-oriented education system has been identified as a barrier to the teaching of communicative language" (H. T. Nguyen, Fehring, et al., 2014, p. 32). If speaking is not included in examinations, neither teachers nor students are motivated to teach and learn speaking skills (Chen & Goh, 2011). The reason for excluding speaking tests has been cited as: "speaking tests cost time and money" (H. T. Nguyen, Fehring, et al., 2014, p. 36), and as a result, students have not had opportunities to practise their speaking skills. The test design and students' desire to pass "tie the teacher to the textbook provided" and students tend to learn passively (Tran, 2013, p. 143). This places a huge strain on teachers who have to juggle the conflicting demands of communicative teaching and preparing students for exams.

English education in Vietnam has been criticised for a lack of standard measurement and effective method for testing speaking (Hoang, 2008). English teachers blame the shortage of interactive activities in classrooms on time limitations and test design. They realise that "the current test design may negate efforts to renew teaching methods, but they just 'go with the flow' because they know that change requires time and commitment. The current teaching style and class organisation invalidate students' efforts, and reduces their motivation and hope" (Tran, 2013, p. 143). Learning for exams deters students from learning communicatively and drives a narrow focus on grammar and reading.

In summary, the importance of English education at tertiary level has been recognised by the Vietnamese government, the Ministry of Education and Training, teachers and students. However, the quality of English teaching and learning at universities is still poor and there has been little improvement in students' English proficiency. Many factors have contributed to this situation, including an imbalance in the assessment

processes for the four English language skills and the absence of speaking tests in universities. It is therefore not surprising that teachers and students have been discouraged from teaching and learning English communication skills.

Challenges of EFL Speaking Assessment

Good English speaking ability has increasingly become a desirable skill and source of cultural capital in workplaces and educational institutions (Isaacs, 2016). The increased emphasis on second or foreign language speaking skills is essential for successful interaction in workplaces (Derwing & Munro, 2009), integration into society, securing employment, overcoming language barriers, performing academic tasks, and effective intercultural communication (Isaacs, 2013). However, the theory and practice of assessing English as a foreign language are misaligned and place greater emphasis on normative and formal aspects of language, such as grammar, pronunciation and spelling, than on the functional aspects, i.e., communication skills (Flores, 2016). Chen and Goh (2011) investigated the obstacles encountered by EFL teachers of spoken English at Chinese universities. In addition to large class sizes, inadequate teaching resources, and teachers' low self-efficacy and poor pedagogical knowledge of spoken English, the authors identified a lack of spoken English tests as one of the impediments.

Although spoken English tests were included in the programs of some universities, "it is only an optional test, which leads to a misconception that oral skills are less important than the other skills" (Chen & Goh, 2011, p. 16). Aleksandrzak (2011) argued that speaking should be included in language tests because it is generally considered to be the most important language skill. The author claimed that testing English oral proficiency will guarantee teachers and students spend more time practising, teaching and learning speaking, which he observed as a washback effect on pedagogy in his study. According to Chen and Goh (2011, p. 10), "oral English is not given adequate attention in the syllabus and the testing system and this gives rise to a negative washback effect on oral English teaching". Aleksandrzak (2011) also argued that speaking tests ensure fairness to all students by allowing those who are better at speaking than writing to demonstrate their proficiency (2011).

Nevertheless, "the problems encountered with speaking tests from the early days have not disappeared" (Fulcher, 2014, p. 1). Testing second language oral proficiency is a complex process and problems could arise at any stage, for example, problems with elicitation techniques, forms of assessment, and test administration (Aleksandrzak,

2011). It is also difficult to design valid and reliable speaking tests, because speaking is not easy to assess quickly and objectively. Moreover, "many institutions have made significant investments in the technical infrastructure to support assessment and feedback but this is not yet delivering resource efficiencies due to localised variations in underlying processes" (Ferrell, 2012, p. 3). Some authors view the problem with English speaking tests as the lack of efficient and effective assessment instruments (X. Zheng & Davison, 2008), and the question "What is the most reliable form of speaking assessment?" still needs to be answered.

In Vietnam, MOET provides teachers with training courses in Communicative Language Teaching (CLT), but school examinations focus mainly on vocabulary, grammatical structures and reading (Le, 2013). The assessment of listening and speaking carry little weight in English assessment practice. Although there has been a significant emphasis on CLT to improve students' communication skills, English speaking tests are still not included in the English curriculum of some universities in Vietnam. H. T. Nguyen, Warren, et al. (2014, p. 42) asserted "the exclusion of the speaking component in the tests is the primary reason hindering the teaching of students' English speaking and communication". This disadvantage has led to low motivation for teaching and learning English speaking, and ultimately, shortcomings in students' English communication skills.

In Vietnam, English speaking is not included in achievement tests for non-English major courses; and in English major courses, they are included in summative exams. English speaking assessment has been criticised for being subjective and unreliable, as well as time-consuming (Biggs, 2011). Real-time assessment of speaking competencies without digital recordings of student performances have contributed to this problem. There are no records of students' presentations for later review, standardisation or reflection. Moreover, the lack of qualified English teachers results in little interaction when grading student achievement, because they are graded individually (Allal, 2013).

Thus, there is a critical need to find an effective and manageable way to assess English speaking skills reliably in Vietnam. A digital testing method that allows multiple markers to access and mark student performances presents a viable solution to current problems relating to test reliability, objectivity and fairness.

Context of the Study

Data were collected from EFL teachers and students at FPT University in Vietnam, a mainly technical university. It was equipped with modern learning and teaching facilities and all classrooms had projectors, speakers, and Wi-Fi connection. First-year students were provided with a laptop by the university, which they used for studying and taking tests. Most of the communication among teachers and students was via email, the CMS (Content Management System - a university intranet) and other social networks.

FPT University provided training in three main academic areas: Software Engineering, Business Administration, and Graphics Design. According to its mission, objectives and education strategy, English was an integral part of the curriculum and a primary focus of the educational programs. Although FPT students did not major in English, the four English language skills were equally included in all achievement tests, which made this university an ideal context for this study.

Before commencing at FPT University, students had to sit an English placement test. Based on the results, they were grouped into classes aligned with their English competency levels. In their first year at university, students attended English lessons every day of the week. Once they'd completed the highest level of Basic English Education (level five), equal to level C1 in the Common European Framework of Reference for Languages (CEFR) or the band score of 7 in the International English Language Testing System (IELTS), they commenced studying their major subjects. In the ensuing years, they continued to learn English, but focused on Academic Writing and English for Business in fewer lessons per week.

FPT University was selected for this research for two main reasons. First, English speaking was included in achievement tests for all non-English major students at all levels. The findings from this sample can therefore be generalised across a significant number of universities where English is not taught as a major subject. Second, since the study experimented with a digital assessment method for EFL speaking skills, the university had to meet certain basic ICT conditions. Since FPT University possessed modern ICT facilities and its teachers and students enjoyed high levels of ICT competence, it was an ideal location for this research. Last but not least, FPT University was the researcher's previous workplace, which afforded her some advantages with the recruitment of research participants.

Rationale for the Study

Various topics around teaching and learning English in Vietnam have been studied extensively, such as the implementation and introduction of English to primary students in Year 3 by H. T. M. Nguyen (2011) and teaching methodology by Hoa and Tuan (2007). Researchers have examined the benefits of native English speaking teachers over non-native EFL teachers in Vietnam and found a correlation with pronunciation (Canh, 2013; Walkinshaw & Duong, 2012; Walkinshaw & Oanh, 2014), but there are no studies that investigate how to improve the overall quality of English speaking assessment in Vietnam. Moreover, little attention has been paid to the integration of ICT in assessing students' English speaking skills, and few studies have been completed on the topic of using digital representation for assessment of EFL communication skills in Vietnam.

Digital presentations for performance assessments have previously been examined in the context of high-stakes summative tests and examinations in four different senior secondary subjects, namely, Engineering Studies (Williams, 2013), Applied Information Technology (Newhouse, 2013), Italian (Cooper, 2013) and Physical Education Studies (Penney & Jones, 2013) in Western Australia. Collectively, these studies showed that digital technologies enhanced the reliability, authenticity, and manageability of academic subjects assessment (Newhouse, 2011). As far as the researcher is aware, the feasibility of using digital representation for assessing students' English speaking performance has not been explored in the literature.

Another reason for undertaking this study was that paper-based assessments of English competency cannot meaningfully and adequately assess performance. Digital representation of assessment can capture complexities in performances that would otherwise not be available to facilitate marking and review. In addition, digital assessment allows records of performances to be retained for later review and reflection, and provides access to multiple markers and collaboration, thereby enhancing reliability and validity.

Purpose of the Study

This study examined the feasibility of applying digital representation as an assessment method to EFL speaking skills in universities in Vietnam, explored across four different dimensions: technology, functionality, pedagogy and manageability. It also brought to the fore the advantages and disadvantages of the digital testing method in the particular

context of English education in Vietnamese universities. Educational organisations are urged to consider the use of digital representation for EFL speaking assessments in particular and for other subjects more broadly, to improve reliability and fairness.

The intention behind the study was to fill the gaps between how English language is taught, what English skills are being learnt and what is being assessed in the current testing methods in Vietnam (Hoang, 2010). It was specifically designed to address the exposed misalignment between the standards expected to be mastered by students and those that were actually being taught, learnt and assessed (Le, 2013). The inclusion of EFL speaking in important language tests and examinations at universities, was also placed under the spotlight.

Previous research found that "academic staff have too few opportunities to gain awareness of different approaches to/forms of assessment because of insufficient time and a lack of opportunities to share new practices" (Ferrell, 2012, p. 3). This study provided teachers with an alternative testing method that allowed them to reflect on the differences between the conventional method and the digital one.

Significance of the Study

The research contributes to the paucity of literature on improving the process of conducting EFL oral proficiency assessments in Vietnam. It addresses the poor reliability of current English speaking assessment methods, and it is hoped, will encourage tertiary institutions to add a speaking component to English achievement tests and examinations. In addition, teachers and students are likely to be more motivated to teach and learn English communication skills, lending support to the National Foreign Languages Project 2020 (NFLP/2020 project) (MOET, 2008) and others, including the Decision of Adjustment and Supplementation of the National Foreign Languages Project 2020 for the period 2017-2025 (MOET, 2017). The Decision emphasises the importance of language assessment for improving language teaching and learning and recommends enhanced assessment methods and integrated ICT.

The acquisition of speaking skills for gainful employment and full participation in academe, international integration and exchanges holds the promise of a positive outcome for students in the form of a pathway to higher education, professions and careers. To this end, the study includes recommendations for assessment policies, such as the inclusion of English speaking assessment in high-stakes examinations. Such a move is likely to have a motivating impact on teachers and students' attitudes that will

translate into higher numbers of quality graduates from tertiary institutions. The current study can also serve as a reference for other countries where English is taught and assessed as a foreign language.

This thesis contributes to the existing body of knowledge on the integration of ICT in English speaking assessment. The investigation has generated valuable new knowledge about digital performance testing and will be of interest to students, teachers, language assessors, and the research community.

Scope of the Study

The study was undertaken in two phases. Phase 1 involved exploring student and teacher perceptions about the implementation of computer-assisted EFL speaking assessment and their willingness to trial a speaking test. In Phase 2 the study focused on the assessment process using video recordings of student speaking performances. The recordings were uploaded to the internet together with the markings embedded in Oral Video Assessment application (OVA App) designed using FileMaker Pro. The OVA App was custom designed by Dr Alistair Campbell at the Centre for Schooling and Learning Technologies (CSaLT), School of Education, Edith Cowan University, Western Australia, and adapted for the context of FPT University. Teachers logged into the online database of student performances to complete their marking, after which correlations were examined between the digitally and conventionally marked outcomes.

The feasibility of digital representation for assessment of EFL speaking at tertiary level in Vietnam was investigated through the lens of Kimbell et al.'s (2007) feasibility analysis framework and the four dimensions of technology, manageability, functionality and pedagogy. The functional dimension was a combination of assessment qualities, i.e., fairness, reliability and validity.

Although listening skills contribute to students' speaking performance, they were not included in the assessment criteria of the current study. Also, although students were provided with speaking questions on paper that required them to read and understand the questions, reading skills were not assessed either. The study was limited only to the assessment of students' speaking competence, based on a marking key that was adapted from one being used at FPT university and the public version of the IELTS speaking marking key.

While the study was conducted at one particular university in Vietnam, the context was sufficiently typical for the findings to be generalisable to the other educational

institutions in Vietnam and beyond, where similar environments for teaching, learning and assessing English as a foreign language occur.

Research Questions

The research was borne out of concern for the issues associated with the assessment of EFL speaking in tertiary education in Vietnam, as frequently referenced in the literature. Currently, EFL speaking is included in achievement tests at few universities in Vietnam, ones where English is taught and learnt as a major subject. The vast majority of universities and colleges do not include English speaking in tests and examinations for several reasons. First, English speaking tests are time-consuming and costly. Most universities do not have sufficient resources, including English teachers and time, to undertake speaking tests with a large cohort of students. Second, the quality of current English speaking tests is questionable, due to high levels of subjectivity and individual judgment by one person or another. Reliability of the current speaking test method is also contestable, because they are conducted in the form of face-to-face interviews and leave no evidence of student performances for later marking and review. Due to a scarcity of teachers tests are marked by one person only and recordings do not exist for other teachers to review.

These issues have persisted for a considerable time and no solutions have yet been found. In Western Australia, a group of researchers at CSaLT Centre, School of Education, Edith Cowan University, completed a series of research projects using digital representation to assess student performances in certain subjects with the aim of improving the quality of the process. The method proved suitable for assessing performances such as dance and Italian speaking.

Digital representation is considered cost-effective, because it does not involve huge sums of money associated with technologies, storage and internet bandwidth. The method retains student performances, delivers them to the internet, and provides easy access for multiple teachers and assessors. In the context of digital assessment and English education in Vietnam, the main research question was therefore:

How feasible is digital representation for summative assessment of EFL speaking performance in Vietnam?

The main research question was underpinned by three subquestions:

1. What are teacher and student perceptions of computer-assisted EFL speaking assessment?
2. What is the feasibility of digital representation of student performances for English speaking assessment in terms of functionality, manageability, pedagogy, and technology?
3. What are the benefits and limitations of digital representation of students' performance for summative English speaking assessment in Vietnam?

Subquestion 1

What are teacher and student perceptions of computer-assisted EFL speaking assessment?

As previously mentioned, face-to-face interviews have traditionally been used to assess students' English speaking competence, and the teachers and students were familiar with this mode of testing. To introduce a new method that used modern technologies for assessing English speaking required certain preconditions, notably teachers' and students' competence in information technology, their general knowledge of computer-assisted language assessment (CALA), and in particular, their willingness to trial a digital speaking test. Other information about school resources and demographics, such as teachers' experience and students' English levels, was also needed for the study.
Davis, Bagozzi, and Warshaw's (1989) technology acceptance model was adopted to investigate teachers' acceptance of computer-assisted language assessment. Teachers' beliefs and attitudes are further discussed in relation to their willingness to participate in a trial of the new testing method. Data on students' perceptions of computer-assisted English speaking assessment (CAESA) were collected and analysed using descriptive statistics and qualitative theme coding. Teachers' and students' attitudes towards the trial were also compared.

Subquestion 1 of the study was addressed by the following three questions:

1. What language testing techniques are currently used in Vietnam?
2. What are teachers' and students' views of computer-assisted assessment (CAA)?
3. Do teachers and students show an attitude of willingness toward the introduction of a computer-assisted assessment trial?

Subquestion 2

What is the feasibility of digital representation of student performances for English speaking assessment in terms of functionality, manageability, pedagogy, and technology?

The feasibility of implementing digital representation for EFL speaking assessment was investigated across four different dimensions: technology, manageability, functionality and pedagogy, adapted from the feasibility analysis framework of Kimbell et al. (2007).

In terms of technology, the extent to which existing technical facilities at FPT University could be adapted, were examined. Students and teachers provided feedback via surveys, and as the main stakeholders in the assessment process, teachers expressed their views about adapting the facilities to accommodate the new technology. This dimension also covered the IT competence of teachers and students to determine whether they could manage the technology.

The manageability dimension covered administration of the assessments, including collection, storage and distribution of students' work and results, as clarified in the description of the OVA App. Since this was the first study to use the OVA App, these aspects were managed by the researcher and her supervisors. Issues regarding feasibility of the new assessment method in normal classrooms and training for teachers and students were also included in the investigation.

Functionality referred to the validity and reliability of the digital assessment method, addressed by a correlation coefficient analysis of student results, teacher surveys and interviews.

The pedagogy dimension looked at how digital assessment supported and enhanced EFL teaching and learning, and whether it enhanced reliability and fairness. The study explored the ability of digital assessment to encourage teachers and students to reflect on their delivery and performance respectively. In addition, the pedagogy dimension examined whether digital assessment addressed any weaknesses in current teaching, learning and speaking practices.

Subquestion 3

What are the benefits and limitations of digital representation of students' performance for summative English speaking assessment in Vietnam?

The benefits and limitations of digital assessment were investigated via teacher and student perceptions in surveys and interviews. Comparing and contrasting the new and existing testing method helped to identify the benefits and limitations of the new model and how they could be addressed for large-scale implementation. The answer to this subquestion was intended as an indicator for recommending implementation of digital EFL speaking assessments in the future.

The study made use of the following innovations:

- Students' EFL speaking performances were captured on video and stored in digital files.
- The digital records were placed in an online repository for easy access by multiple markers.

Thesis Organisation

The thesis is organised into seven chapters. Chapter 1, the Introduction, provides an overview of the study, the background to the research, the context, rationale, purpose, significance, and scope of the study. The research questions are also listed.

Chapter 2, the Literature Review, presents a critical review of the relevant literature in relation to the theoretical background and conceptual framework of the study. It covers two main areas, viz., English Education and Educational Assessment.

Chapter 3, Methodology, outlines the methods adopted to collect data for the study in order to answer the research questions. Mixed method and case study approaches are reviewed and the research design presented.

Chapter 4 gives an analysis of the Phase 1 data and findings, the preliminary phase of the study. During this phase, data were collected on the ICT competence of teachers and students, their CALA knowledge, and their willingness to participate in the digital assessment trial conducted in Phase 2.

Chapter 5 presents the Phase 2 data analysis and findings investigating the feasibility dimensions of DMOVA and the benefits and limitations of its implementation. Chapter 6 contains a discussion of the findings based on the conceptual framework and research questions, and Chapter 7 concludes the study and presents recommendations for practice, policy and further research.

CHAPTER 2
LITERATURE REVIEW

This, the literature review chapter, focuses on English education and educational assessment. English education covers second language acquisition and ESL/EFL teaching, including the use of technologies in English teaching. It hones in on teaching and assessment of English speaking, for which marking methods are an indispensable part of assessment. The second aspect of the literature review, education assessment, covers different assessment types and their characteristics, assessment tasks, task assessment and stakeholders. Performance assessment, second-language assessment, computer-assisted language assessment, and the use of digital representation in assessment are included. These aspects formed the theoretical background and conceptual framework for the research.

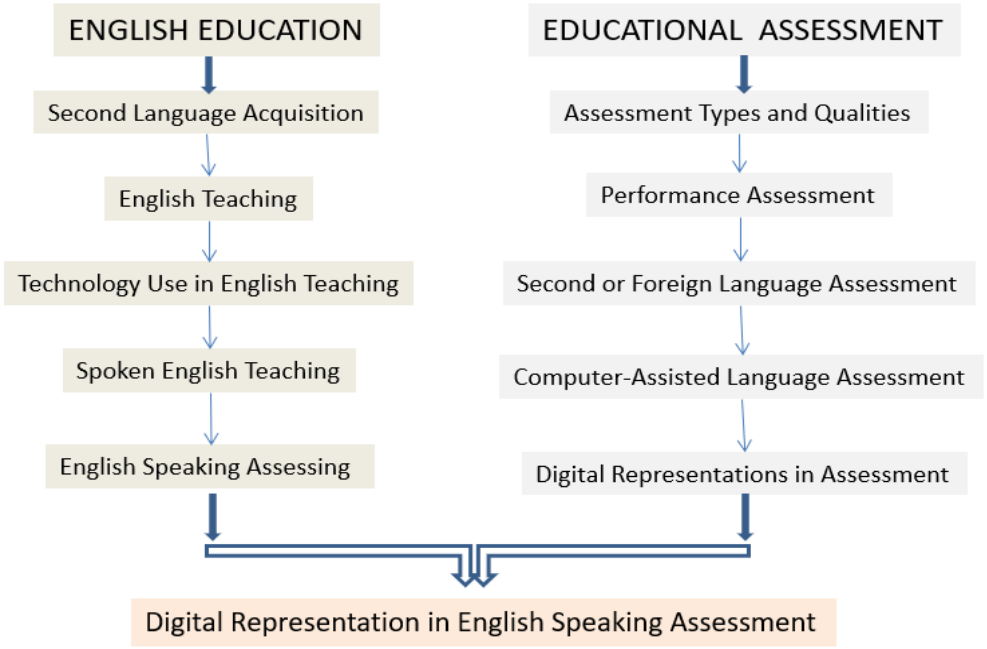


Figure 2.1 Diagrammatic Overview of the Literature Review.

English Education

Second Language Acquisition (SLA)

Language is undeniably one of the most unique human abilities (Ortega, 2014, p. 1). People normally use the language they were born and grew up with, namely their mother tongue, to communicate with others and the world. Some people grow up speaking more than one language in their homes (Harmer, 2014). However, under some

circumstances and for different reasons, people need to learn a second language that is different from their first, and which they are required to communicate in. First language acquisition, believed to go hand in hand with mental and social development, is different from second language acquisition (Cook, 2016). How a second language is acquired and the factors that assist second language acquisition have been widely studied and numerous theories posited by different linguists and researchers around the world. The following table provides a list of different theories and hypotheses proposed since the beginning of the study of SLA. These theories and methods have influenced second language education and generated much debate among educators and researchers.

Table 2.1

Theories and Hypotheses of Second Language Acquisition

Time periods	**1940s - 1950s**	**1960s - 1970s**	**1980s - present**
Theories and Methods	Behaviourism, S-R-R (Stimulus, Response, and Reinforcement)	Nativism. Universal Grammar, LAD (Language Acquisition Device)	Social Interactionism, Output Hypothesis
Authors	Skinner	Chomsky, Krashen	Vygotsky, Swain

Adapted from Malone (2012)

Ellis (2010) maintained two main factors addressed the question *How do learners acquire a second language*? The author envisioned a conceptual framework for SLA research, whereby researchers could identify the external factors that contribute to acquiring a second language, such as the social situation in which the learning takes place, language input, and learners' language production or output. In addition, internal factors, such as mental processes, existing knowledge of mother tongues and learning strategies, as well as universal characteristics of languages could be examined to see what and how they contributed to SLA. Ellis (2010) emphasised that both internal and external factors, and the interrelationship between them, should be considered in language acquisition.

SLA theories belong to one of three different schools of thought: (a) behaviourist; (b) nativist; or (c) interactionist. The theory of behaviourism, proposed by Skinner, rose to popularity between the 1930s and 1950s, and purports that learning occurs by generating responses to positive and negative stimuli and reinforcement. According to this theory, reward encourages positive behaviour and punishment prevents negative behaviour. The disadvantage of this theory is that it turns out passive students because it is essentially a teacher-centred approach.

At the other end of the spectrum, Noam Chomsky argued that children are born with an innate understanding of grammar and syntax, which explains their ability to rapidly acquire language. Chomsky developed the concept of language acquisition device or LAD in the 1960s (Kozulin, Gindis, Ageyev, & Miller), believed to be imprinted in children's brains, readying them for taking on a new language. Chomsky also developed the theory of universal grammar, claiming that all human languages are built on common rules and children are born with these sets of rules in their brains. They pick up and copy the language they hear while learning and use LAD to generate appropriate language patterns. In contrast to behaviourism where learners generate language patterns based on external stimuli and conditions, LAD encourages learners to produce new patterns without any formal instruction. Innatist perspectives are linked to the critical period hypothesis, asserting that knowledge can be acquired more rapidly at certain specific times of life (Lightbown & Spada, 2013). Chomsky encountered criticism for his heavy emphasis on grammatical rules and ignoring the role of interaction in learning a new language. While Chomsky's theory is relevant, it is insufficient for describing the complete process of language acquisition.

Cognitive theory was put forward by Piaget (1976) to explain how children acquire knowledge, after concluding that biological maturation and interaction with the environment determine the process of children's knowledge acquisition. The author determined that language acquisition occurs when children interact with the environment and construct learning; a language learning process where students are central and contribute actively. However, the role of social setting and culture are not mentioned in Piaget's theory as contributing factors to children's knowledge acquisition (McLeod, 2018).

The important role of social interaction in cognitive development was embodied in Vygotsky's sociocultural theory, whereby thought is viewed as internalised speech that emerges during social interaction. Social interaction improves language and thinking abilities, and constructs learners' knowledge (Lightbown & Spada, 2013, p. 37). Vygotsky claimed that a child acquires knowledge through interacting with people, internalising and intermingling the knowledge with personal values (Turuk, 2008). Moreover, "the theory asserts that learning is a collaborative achievement and not an isolated individual's effort, where the learner works unassisted and unmediated" (Turuk, 2008, p. 258). Vygotsky put forward the scaffolding theory to describe a process whereby teachers provide students with guidance and modelling, subsequently

stepping back and lending support when needed. With the teacher's guidance, learners move from understanding to independent learning and acquiring knowledge for themselves. Vygotsky identified the importance of conversations between children and adults and amongst themselves, claiming they contained the origins of both thought and language and provided children with scaffolding to structure and acquire knowledge (Lightbown & Spada, 2013). Scaffolding theory is important for encouraging students to learn actively and independently and allows teachers to push students beyond their current levels of competency (Hammond & Gibbons, 2005).

Well-known linguist, Krashen (1982), claimed that second language acquisition comes from communicative and comprehensible input, and SLA is more efficiently achieved by learners who possess high self-motivation, self-confidence and less anxiety. Hence, learners should be provided with large amounts of comprehensible input in a relaxed setting (Harmer, 2014), particularly for mastering writing. The author hypothesised that sufficient input is necessary to master spontaneous communication, in varying amounts and types according to the learning objectives and mode of interaction. Although comprehensible input is essential for SLA, it is not sufficient on its own. Swain (2005) stated that output is not simply the product of language learning but a part of learning, and proposed the output hypothesis, with three distinct functions. The "noticing" function occurs when learners identify a gap in their linguistic knowledge and attempt to fill the gap by communicating. The "testing" function describes learners using the target language to communicate, making mistakes and receiving feedback that helps them to understand the language. The "reflective" function explains learning a target language through the influence of teachers' and learners' conversational partners. Swain's hypothesis emphasises the importance of language production, including writing and speaking, requiring learners to use the target language appropriately to successfully construct second language production (Ellis, 2010).

In SLA, groupwork can be effective for increasing language practice and improving the quality of student talks (Ellis, 2010). Interaction in small groups promotes a positive atmosphere and motivates learners, while in larger classes, groupwork maximises student participation (Harmer, 2014). Porter (1986) cautioned that groupwork is less collaborative with learners who possess different levels of language proficiency, because more competent individuals will naturally be more gregarious than their less competent counterparts.

This review of SLA literature showed that Vygotsky's sociocultural theory and Swain's output hypothesis support the acquisition of language by encouraging interaction and communication among language learners. Therefore, they were adopted in this study to provide background and a theoretical framework for analysis and discussion of the pedagogical impacts.

English Teaching

Teaching English is a huge industry around the world, comprising millions of students variously described as learners of English as a Second Language (ESL) or English as a Foreign Language. Harmer (2014) defined ESL learners as people who migrate to English-speaking countries and need to learn the language to communicate with the locals. EFL learners are those who study English in their own countries without the same priorities and opportunities as ESL learners. Another branch of English teaching is known as English for Specific Purposes (ESP), such as for science and technology or law. There is also a branch of English teaching called English as an Additional Language (EAL), which refers to students who live in countries where English is the predominant native language but for whom English isn't their first language.

Throughout the history of language teaching, different agendas and modes of teaching have been prioritised, and over time, language teaching methods have shifted from grammar-translation to communicative language teaching (J. Richards & Rodgers, 2014). Despite the introduction of new teaching methods, as shown in Table 2.2, "there is not one single best method for everyone in all contexts, and … no one teaching method is inherently superior to the others" (Alemi & Tavakoli, 2016, p. 1). Every method is most effective when it is used appropriately for learners' specific purposes, learning style and context.

The grammar-translation method enjoyed a significant period of influence during the 20th century. It refers to a method of explaining grammatical rules and then applying the knowledge by translating sentences and texts into the target language. Reading and writing are the main foci of this teaching approach, with speaking and listening receiving little or no attention. There is an emphasis on accuracy, and the students' first language is the medium of instruction in the classroom (J. Richards & Rodgers, 2014). Translation, focused on acquiring lists of grammatical rules and vocabulary, is widely considered to have the least effect on EFL learning (Cook, 2016). Nevertheless, the grammar-translation method is still effective in contexts where accuracy is the English learning objective (S. Chang, 2011).

Table 2.2

Language Teaching Methods

Years	Language Teaching Methods
1900-1970	Grammar Translation
1910 - 1950	Direct Method
1940 - 1985	Audio-lingual Method
1975 - 2000	Suggestopedia
1970 - 1980	The Silent Way Community Language Learning
1970 - now	Total Physical Response
1980 - now	Communicative Language Teaching
2000 - now	Principled Eclecticism

Adapted from A. Taylor (2015).

Similar to learning the mother tongue, naturalistic principles of language learning emerged in response to the shortcomings of the grammar-translation method. They were first applied by Sauveur (1826-1907) in his private language school in Boston. Referred to as the "direct method", the principles guide teachers to use the target language extensively for instruction without translating. According to this method, learners acquire language by associating meaning from the mother tongue and applying it directly to the target language (A. Taylor, 2015). Although the direct method was effective in enhancing language learners' communication skills, it was criticised for lacking a methodological basis (J. Richards & Rodgers, 2014).

The audiolingual method, based on Skinner's behaviourism theory, was popular between the 1950s and 1970s. This teaching process focused on drills to form habits, imitating teachers' utterances, and students' pronunciation to gain mastery based on memorisation (Cook, 2016; Harmer, 2014; Savignon, 2017). Although the audiolingual method was effective in forming habits, "much audiolingual teaching stayed at the sentence level, and there was little placing of language in any kind of real-life context" (Harmer, 2014, p. 57). This method has been criticised for not developing long-term communicative ability in language learners (Savignon, 2017).

Prior to communicative language teaching (CLT), many other language teaching methods were proposed, including the Silent Way, Total Physical Response, Community Language Learning, and Suggestopedia. Task-based language teaching and

content-based language teaching originated from sociocultural theory and viewed language acquisition as constructed through social interaction (J. Richards & Rodgers, 2014). Between the 1970s and 1985, these methods were an attempt to improve language teaching, a purpose they served with worthy attention. Task-based language teaching is still used today.

Linguists and language teachers criticised the grammar-translation and audiolingual methods for their incapacity to provide learners with communicative opportunities (Savignon, 2017), giving rise to an alternative teaching method that fosters communicative competence. In reality, "most English teachers in the world today would say that they teach communicatively" (Harmer, 2014, p. 57). Communicative language teaching (CLT) proposes that language be taught holistically, through meaningful communication and interaction. Although CLT is interpreted differently by different people (Harmer, 2014), the method focuses on enhancing learners' communicative competence both in the classroom and real-life contexts (Jackman, 2016). CLT activities include role play, games, debates, and discussions. These activities are encouraged in the classroom via social interaction, where learners are motivated to share their opinions in pairs or groups (Loumbourdi, 2018).

CLT textbooks were a shift away from current teaching approaches, focusing on language skills training and communicative activities. However, "tests continued to focus on discrete language items" (Harmer, 2014, p. 58), making it difficult for teachers to convince students of the importance of communication. At the same time, teachers were challenged to be communicative in their English teaching practice.

The CLT approach has been proven to enhance students' communication skills by exposing them to authentic speaking situations, where they are able to express themselves and learn appropriate social and cultural rules for different social circumstances (Kayi, 2012). It was derived from interactional second language acquisition theory that focuses on learners' negotiation of meaning or modifying the input and feedback they receive from interaction with others to support understanding and learning (J. Richards & Rodgers, 2014). CLT has gained popularity over other teaching approaches for its capacity to develop the ability of learners to use English for communication from the perspective that "What people want to do through language is more important than the mastery of language as an unapplied system" (Thornbury (2016, p. 225). However, in order to get the best from CLT, Thornbury recommended that assessment should be compatible with the communicative language teaching

method, and it should be applied appropriately and flexibly in diverse contexts of English teaching, including teaching and learning English as a foreign language.

In Vietnam, CLT has been the principal EFL teaching method for improving students' English communication skills since it was first introduced in the early 1990s (Ngoc & Iwashita, 2012). In spite of early adoption in the school system, the quality of EFL teaching and learning in Vietnam is still below expectations (Hoang, 2010; Tran, 2013). Previous studies have shown that CLT was not properly and effectively implemented due to insufficient time for communicative activities in classrooms (H. T. Nguyen, Warren, et al., 2014). In addition, crowded classrooms have diminished speaking opportunities and communication practice for students. Test-oriented teaching styles remain popular and teachers spend a significant amount of time teaching and explaining grammatical rules that could be reviewed by students at home. Nguyen, Warren, et al. (2014) recommended that EFL assessment should cover the four language skills equally. Hiep (2007) encountered numerous difficulties implementing CLT in a Vietnamese context, even though the teachers willingly embraced basic CLT principles in their teaching practice. Thornbury (2016) proposed that CLT in Vietnam be adopted flexibly, together with transformative ways of testing English, to ensure that the goals of communicative English teaching and learning are achieved and English communicative competencies enhanced, as directed in the National Foreign Languages Project 2020 (NFLP/ 2020 project).

Use of Technology in English Teaching

The adoption of technology in teaching, particularly language teaching, has been extensively and intensively researched with the aim of enhancing effectiveness. English language teaching is no exception. Although the grammar-translation method was the most influential teaching style at the beginning of the 20th century, audio-visual technologies were introduced into classrooms by teachers of Latin and German to help students practise speaking and listen to the accents of native speakers (Otto, 2017).

Over the decades, teaching methods have changed with the tide to incorporate technological advances and adapt to the growing numbers of students in and of the digital generation. Integrating information and communication technology (Reynolds, Livingston, Willson, & Willson, 2010) into teaching and learning brought about significant educational benefits and positively changed the learning environment (Ahn & Lee, 2016; Floris, 2014). Many computer-assisted teaching and computer-assisted

language learning (CALL) methods have been adopted to facilitate teaching and increase the language competence of learners, including blended learning, first introduced in 1998. These methods were aimed at enhancing the quality of teaching and learning and promoting engagement and motivation. Today, the internet and multimedia offer language learners more opportunities to acquire new knowledge, practise their language skills, and share learning experiences, with abundant benefits for both learners and teachers. (Floris, 2014; Houcine, 2011)

Rusanganwa (2013) asserted that the use of technologies in education facilitates teaching and learning. In many ways, technology now plays an important role in language teaching classrooms, as reported by Stanley (2013) and Padurean and Margan (2009). Computers serve as teachers, testers, and communication facilitators, and provide tools and data sources that create appealing and authentic learning environments with texts, graphics, sound, animation, and video all linked together.

ICT has also been found to advance student-centred learning (Mullamaa, 2010), increase student motivation (Facer & Owen, 2005; Stockwell, 2013), interaction and collaboration via web-based learning environments (Pais Marden & Herrington, 2011, 2020), and provide access to databases, PowerPoint presentations, and online dictionaries. Language skills are enhanced through interaction (Alsied & Pathan, 2013), so the more interaction language learners are exposed to, the more proficient their language becomes (Morozova, 2013). Fitzpatrick, Davidson, Davies, Diakite, and Lund (2004) concluded that digital media fostered closer interaction between teachers and students. Furthermore, a web-based learning environment creates an online community of language learners who interact socially and learn collaboratively with native speakers through authentic activities (Pais Marden & Herrington, 2020). ICT helps open up new spaces and opportunities for communication, bringing about a “youth culture of hybrid language practices” (Fitzpatrick et al., 2004, p. 28).

ICT also contributes to language learning by providing access to authentic materials and communication via video conferencing. Multimedia presentation software allows students to practise their language skills; while digital video provides feedback on students’ language performance for self-critique, teacher and peer evaluation. Students can work at their own pace while their autonomy is supported (Kirkgoz, 2011; Klimova, 2012; Maryam, Ahmad, Elham, & Nasrin, 2013). In a study by Maryam et al. (2013), ICT proved to assist teachers develop highly interactive classes and adopt new techniques for enhancing learners’ communicative competence.

In spite of its significant benefits, the use of technology in language teaching and learning poses a challenge for students who have low levels of ICT proficiency and may result in widening gaps between teachers and learners (Uzunboylu & Tuncay, 2010). It is also possible for there to be a misalignment between teachers' interest in adopting ICT and the extent to which they integrate ICT into their practice (Wang, 2014). While many express a positive attitude towards the use of ICT, some experience anxiety and a lack of confidence due to the absence of proper training, insufficient technical knowledge and the spectre of equipment malfunctions.

Integrating ICT into English language teaching poses some challenges in terms of implementation, and requires ongoing training, technical support, and an awareness of pedagogical philosophy (Hadi & Zeinab, 2012). Similarly, when the internet - a powerful resource for English language teaching - is incorporated into the program, it is necessary to redesign the curriculum and pedagogical practices. Hu and McGrath (2012) indicated that teachers and students were overwhelmed by e-materials and blamed an overly zealous focus on technological presentations and adaptations for the lack of teacher-student interaction in the classroom. In their case study in China, Hu and McGrath (2012) identified limitations in the ICT competence levels of most EFL teachers, who mainly used the email, search and download functions to access material on the internet, and PowerPoint for presenting lessons. They needed more training in the use of Web tools and other software to competently and confidently incorporate ICT in their classrooms.

Regardless of the challenges and difficulties, ICT creates an ideal environment for authentic language teaching and learning, unhampered by geographical borders and time zones. Negoescu and Boştină-Bratu (2016) asserted that ICT offers the advantage of interactivity, including interactive applications to language learning and teaching. According to Hu and McGrath (2012), ICT provides rich learning resources with authentic and updated audio and video records – "a reality beyond the classroom walls" (p. 30).

The internet also offers powerful tools and advantages for English language teaching and learning. Zamorshchikova, Egorova, and Popova (2011) stated that "ICT as tools of e-learning in teaching EFL are becoming more widespread in higher educational institutions and are meeting education quality requirements" (p. 75). Notably, ICT opens up opportunities for international and cross-cultural collaborative projects. According to Zamorshchikova et al. (2011), teachers and learners should actively

change their conventional teaching and learning styles to keep up to date with new and effective techniques available to them.

Spoken English Teaching

Speaking is an important language skill that facilitates communication and helps learners acquire proficiency (Bashir, Azeem, & Dogar, 2011; Goh, 2007). Mastery of speaking skills is considered an important measure of knowledge of a particular language. Nazara (2011) argued that the more learners master speaking skills, the more they master that language. Speaking competence requires considerable attention and practice through regular interaction, whereby language learners produce language and receive feedback from listeners (Bashir et al., 2011). The comprehensible output hypothesis, developed by Swain (2005), theorises that second language acquisition takes place when learners become aware of a gap in their linguistic knowledge (in writing or speaking) and try again. Feedback plays an important role in helping learners reflect and improve their linguistic knowledge. The hypothesis supports the idea that the output or language production (speaking and writing) in the target language aids language acquisition.

Hinkel (2017) defined teaching second language speaking skills as helping language learners master specific sets of interactional and communication skills. When learning a second language, learners are required to develop their speech-processing, discourse organisation and oral production skills, including correct grammar, rich vocabulary, accurate pronunciation, and information sequencing (Hinkel, 2017). As a productive skill, speaking is widely believed to be the most important of the four language skills, because it reveals any errors made by the learner (Khamkhien, 2010) and is the main way of communicating and forming relationships with people. However, "for many years, teaching speaking has been undervalued and English language teachers have continued to teach speaking just as a repetition of drills or memorisation of dialogues" (Kayi, 2012, p. 1). Goh (2007) stated:

> Unlike with lessons on reading and writing where the teachers will have a record of performance in the form of written texts, speaking output is transient, with little record of it once the activities are over. Teachers do not have a corpus of learner work which they could evaluate and give feedback on. As a result, problems that learners face when doing speaking activities often go unnoticed or uncorrected (p.1).

The phenomenon of English as a lingua franca (ELF) emerged recently and refers to communication in English between speakers of different first languages (Seidlhofer, 2005, 2013). The majority of English users speak English as a foreign language, and the majority of verbal instructions and interactions in English do not involve any English-native speakers (Seidlhofer, 2005). Therefore, overemphasis on a British-native accent would be inappropriate in non-British settings (Harmer, 2014). For learners who use English as a lingua franca, it is not necessary to achieve native-like competence or sound like native speakers (Kirkpatrick, 2011). Kirkpatrick pointed out that regional or non-native English language teachers, rather than native English teachers, provide students with linguistic norms and models. It is therefore crucial that teachers are tolerant in assessing and providing feedback on the use of non-native pronunciation and expressions (Snow, Kamhi-Stein, & Brinton, 2006).

Throughout the history of language teaching, priorities have shifted away from reading comprehension to oral proficiency and from grammar-translation to communicative language teaching (CLT) methods (J. Richards & Rodgers, 2014). In the Asia-Pacific region, CLT is widely used in English curricula to advance English communication skills (Butler, 2011). However, problems related to teachers' perceptions and beliefs about teaching speaking, curricula, teaching strategies, the lack of qualified English teachers, and assessment policies have resulted in limited adoption of CLT for improving EFL oral proficiency (Al Hosni, 2014; Butler, 2011; Khamkhien, 2010; Khan, Shah, Farid, & Shah, 2016). Khamkhien (2010) and Khan et al. (2016) identified that little time and attention were being paid to teaching EFL speaking compared to reading and writing. EFL teachers mainly focused on students' grammatical competence, pattern drills and memorisation of individual sentences to the exclusion of authentic speaking activities.

First language (L1) interferes with the process of acquiring English and causes mistakes in pronunciation and sentence building. It is difficult for teachers to encourage students to make accurate utterances in authentic settings when English speaking tests do not motivate students to produce natural, authentic output. In such ways, speaking tests undermine positive washback effects on teaching and learning English speaking skills.

In summary, the partial adoption of CLT in English teaching and lack of appropriate assessment policies appear to be the key factors underlying the limited success of teaching and learning EFL speaking skills (Al Hosni, 2014; Kayi, 2012). In fact, "many teachers are familiar with the situation where their own beliefs in CLT, for example, are

at odds with a national exam, which uses an almost exclusively discrete-item indirect testing procedure to measure grammar and vocabulary knowledge" (Harmer, 2014, p. 421). Aleksandrzak (2011) proposed changes in EFL speaking assessment to guarantee teacher and student engagement in practising, teaching and learning English speaking skills in order to ensure fairness for all students, especially those who are better at speaking than writing.

English Speaking Assessment

Assessment Methods

Luoma (2004, p. 1) claimed that "speaking skills are an important part of the curriculum in language teaching and this makes them an important object of assessment as well". English speaking assessment mainly evaluates improvements in students' pronunciation and communication (Khamkhien, 2010), and in many contexts, students' communicative competence is still assessed by means of multiple choice paper-and-pencil tests (Sinwongsuwat, 2012). It is essential for communicative tests to "find out what a learner can "do" with the language, rather than to establish how much of the grammatical/lexical/phonological resources of the language he/she knows" (Morrow, Coombe, Davidson, O'Sullivan, & Stoynoff, 2012, p. 40).

Although "… most language test users really value the ability to communicate in English" (Powers, 2010, p. 3), speaking skills were not tested in certain contexts until fairly recently. For example, TOEFL only included speaking tests in 2005, and TOEIC, in 2006 (Powers, 2010). Speaking tests are still optional for university students in many countries, such as China, Thailand and Vietnam (Hoang, 2010; Khamkhien, 2010; Ying Zheng & Cheng, 2008), and where they are conducted, speaking ability is evaluated against criteria and norm references (Ying Zheng & Cheng, 2008). Tests usually comprise three sections: (a) interaction between test takers and two examiners; (b) group discussion; and (c) further questions and answers to test students' speaking ability.

Speaking is a complicated skill to assess. Brown (2003) advocated for English communicative interaction in speaking tests to be assessed in real contexts of interaction. McNamara (2011, p. 435) claimed "the distinctive character of language testing lies in its combination of two primary fields of expertise: applied linguistics and measurement". English speaking tests need to be valid, which means they must provide

teachers with an accurate picture of what they are intended to evaluate, i.e., students' knowledge and ability to use English (Harmer, 2014).

Testing second language speaking is the youngest sub-field of language testing. Before the First World War, speaking tests received little attention and were avoided because they involved complex problems (see Figure 2.2). In 1913, a sub-test of spoken English was introduced in the form of a Certificate of Proficiency in English in the United Kingdom; marked only for pronunciation using phonetic script, dictation and written answers to questions spoken by examiners. The results from these tests could not provide a true measure of live oral language ability (Fulcher, 2014).

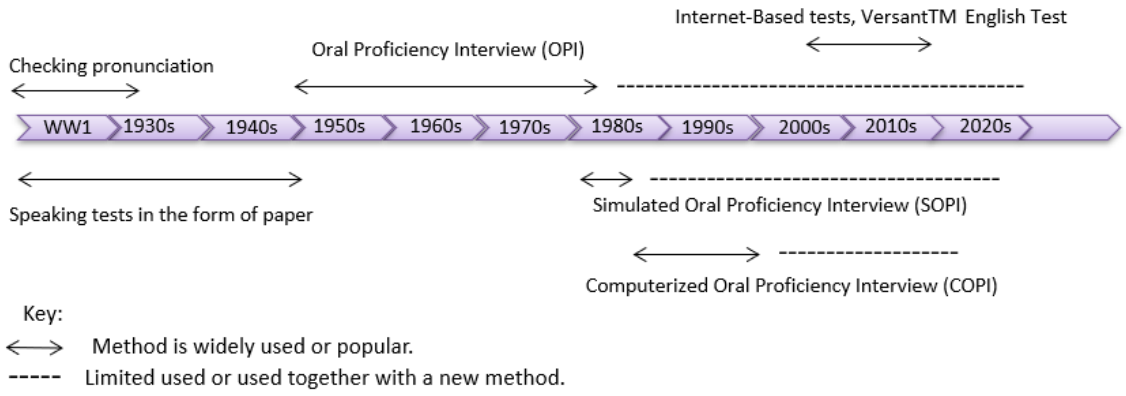


Figure 2.2 Timeline of Second Language speaking assessment methods.

Adapted from Fulcher (2014) and Qian (2009).

In the 1950s, the direct oral testing method was adopted in the United States, where it was named the Oral Proficiency Interview (OPI) or face-to-face oral assessment (Qian, 2009). OPI was conducted by a native interlocutor and a rater, the test comprised of a six-point rating scale across five factors. OPI was considered valid because it simulated conversation and live human interaction, but criticised for subjective judgement, logistical difficulties, inconsistency due to uncontrolled factors, and impracticality for a large number of test takers (Malabonga, Kenyon, & Carpenter, 2005). The variability of human interlocutors also posed a threat to the reliability of assessment (Fulcher, 2014). In addition, OPI was difficult to conduct in remote areas where there was a shortage of certified OPI interviewers (Kenyon & Malabonga, 2001).

The abovementioned issues of reliability and practicality associated with OPI led to development of a semi-direct testing method (Fulcher, 2014), first introduced in the United States in the 1980s, where it was named Simulated Oral Proficiency Interview (SOPI) (Qian, 2009). Tape-mediated SOPI could also be used to test groups of students.

The process entailed using two tape recorders: one containing the master tape that provided instructions and asked the test questions, and the other, the recording of the student's performance (Kenyon & Malabonga, 2001). SOPI was praised for its cost-effectiveness in terms of human resources and logistics, and its ability to enhance reliability and fairness, thanks to removal of the human interlocutor, considered to be the source of errors. However, SOPI also had some disadvantages. In contrast to face-to-face assessment, it failed to generate real-life communication and interaction (Qian, 2009). Nor did it encourage language function, such as negotiating and turn-taking, because the same speaking topics were used with all test takers and the assessment mainly focused on the accuracy of language production (Fulcher, 2014). The Video Oral Communication Instrument (VOCI), developed by The Language Acquisition Resource Center at San Diego State University, was the subsequent version of SOPI and used video recorders instead of tape recorders.

The new generation of SOPI and VOCI was Computerised Oral Proficiency Instrument (COPI), developed in the late 1990s by researchers at the Center for Applied Linguistics in the United States in response to the limitations of SOPI (Kenyon & Malone, 2010; Malabonga et al., 2005). COPI used computer technology and was considered more effective than SOPI, which caused test-takers to be nervous due to a loss of time control. COPI provided test-takers with test samples and a choice of levels: Novice, Intermediate, Advanced, and Superior. It could store a large number of tasks suitable for a large population, generate more authentic speaking tasks, and as the findings showed, encouraged test-takers to perform at their best. Assessors could listen to any part of students' responses several times over and add notes or comments to any part of the test. Kenyon and Malabonga (2001) concluded that COPI fostered positive attitudes toward technology-mediated tests and raised the feasibility of applying computer technology to oral assessment. Nevertheless, COPI was criticised for its inability to replicate the true nature of conversational and interactive face-to-face interviews.

Assessing oral language proficiency online using the internet and other forms of multimedia technology was introduced in the late 20th century (Qian, 2009). At that time, computer-based speaking tests were launched by the Educational Testing Service in the United States. In 2005, a new version of the Test of English as a Foreign Language (TOEFL) was introduced, together with an online speaking test. Since then, improvements and innovation in testing and scoring oral language proficiency have continuously been reported. Developed by the Educational Testing Service,

SpeechRater™ is one example of a system that can automatically score spontaneous non-native speech without human raters. This testing system was used for the TOEFL iBT Practice Online in 2006 (Zechner, Higgins, & Xi, 2007).

Qian (2009) stated:

> Compared with direct testing, semi-direct testing arguably lacks, at least on the surface, sufficient predictive validity because it does not reflect the way most people would communicate in a real workplace, educational or other types of context, except for contexts where technology-enabled communication is heavily used, such as call centers (p. 123).

The direct testing method allowed test takers to communicate with a real interlocutor and use nonverbal expressions to support their verbal communication, as talking to a computer or recorder was criticised for lowering face validity and construct validity compared to real interlocutors (Qian, 2009, p. 123).

Chambers and Ingham (2011) found examiners experienced fewer problems using onscreen marking if they received initial training. In their study, marking was found to be consistent across both modes of paper and onscreen marking. This was a valuable finding and signalled a need for further studies into the feasibility of other forms of marking students' speaking performance than just the face-to-face method.

Feedback in EFL Speaking Assessment

Feedback was defined by Harmer (2014) as teachers' responses, in various ways, to what students say or write. Li and De Luca (2014) decribed assessment feedback as grades and comments that teachers provide in response to work submitted by students for assessment. Assessment feedback should inform learning and justify the teachers' grading, since it contributes to students' learning and future success. According to these authors, constructive feedback must be objective, criteria-referenced, personal and timely, and teachers must make decisions on the kind of feedback to provide and the types of mistakes that need to be corrected. Edge (1989) classified mistakes into three categories: (a) slips, (b) errors, and (c) attempts, with errors the most problematic and needing correction. Harmer (2014) argued it is not necessary to correct every single mistake if it takes time away from other activities. She cautioned against the risk of over-correction when it interrupts the flow of student talks and deters them from engaging in communication and emphasised the need for sensitivity at all stages of correction.

Lynch (1997) suggested that the later feedback is given to learners the better, even after they've finished their presentations. On the other hand, Harmer (2014) argued that on-the-spot feedback is more suitable for activities that focus on accuracy. The recommendation for teachers to give students feedback on the fluency of their communicative speaking activities after they've finished their presentations relies upon memory but is easily solved by writing down the points and comments teachers want to make. Harmer (2014) claimed recording students' performances offers certain advantages. Teachers can identify common mistakes made by more than one student and avoid exposing individual students for their mistakes in front of their classmates. They can also involve their students in peer assessment by asking them to identify their own mistakes, with the purpose of encouraging self-correction and learning.

Marking Methods

Marking is an important part of assessment and needs to be aligned with the curriculum objectives (Herbert, Joyce, & Hassall, 2014). "The grades we give students and the decisions we make about whether they pass or fail coursework and examinations are at the heart of our academic standards" (Bloxham, Boyd, & Orr, 2011, p. 655). Grades must accurately reflect students' effort and improvement (Harmer, 2014). Grades can ultimately encourage or demotivate students, so they should be transparent and based on clear criteria (Dörnyei, 2014).

Analytical marking refers to the process of allocating certain proportions of the marks to different predetermined criteria (Baird, Greatorex, & Bell, 2004; Sadler, 2009). In this way, marking is easier and provides students with detailed feedback and information on their performance (Barkaoui, 2011). The reliability of assessments has been enhanced by the use of rubrics in analytical marking, in turn, supporting learning and instruction (Jonsson & Svingby, 2007). In addition to the use of rubrics, Harlen (2007) recommended internal moderation of teachers' judgments to increase fairness and reliability in summative assessments. However, analytical scoring rubrics have been criticised for being like a checklist and evaluating criteria individually (Moskal, 2000). Raters also tend to be less critical with analytical marking schemes than holistic marking, and therefore, students may be awarded a higher mark for a less deserving performance (Barkaoui, 2011).

A holistic measuring scheme provides a more complete picture of student performances by assessing a collection of criteria (Moskal, 2000). De La Paz (2009) distinguished between the effectiveness of analytical marking that can identify individual students'

strengths and weaknesses, and holistic marking for large-scale assessment. Analytical marking is highly self-consistent, whereas holistic marking leads to higher inter-rater agreement (Barkaoui, 2011). Moskal (2000) argued that both types of marking schemes should be applied to students and assignments and between different markers for maximum consistency.

Moderation "involves teachers of the same subjects or student groups meeting together to align their judgments of particular sets of students' work, representing the 'latest and best' evidence on which the record or report is to be made" (Harlen, 2007, p. 55). Meetings to moderate teachers' judgment are likely to enhance the use of assessment criteria and provide teachers with feedback on their teaching.

Harmer (2014) reported that human markers run the risk of subjectivity because their perceptions of the same students' work are likely to vary. Also, other factors affect the reliability of results assigned by human graders: "assessors have their bad days, too, where they are tired, ill or worried about other matters" (Hartle, 2009, p. 71). Harmer (2014) proposed several ways of enhancing reliability, including training to instil a common understanding of how to score tests and multiple marking of students' work: "two examiners watching an oral test are likely to agree on a more reliable score than one". Harmer (2014, p. 419) also recommended using scales to specify scores in the form of published descriptors, such as the Common European Framework of Reference for Languages (CEFR) and the International English Language Testing System (IELTS), or they could be designed to make the assessment more specific. She argued that scoring should be analytical, particularly for oral assessment, but "a combination of global and analytic scoring gives us the best chance of reliable marking" (p. 420).

Improving the quality of educational assessment seems to be a work in progress for educators, assessors and researchers. Harmer (2014) stated:

> Tests (especially public exams) are, increasingly, administered and graded digitally. Based on extensive trialling and measuring, using experienced scorers coupled with digital analysis, it is claimed that such grading is as reliable as – if not superior to – human marking. And, of course, it is in many ways more efficient, too (p. 418).

In spite of the digital trend, most speaking tests are still conducted face-to-face, their reliability resting on a combination of holistic and analytical assessments. The roles of scorers who mark the tests and interlocutors who guide and provoke conversations need

to be separated. In face-to-face tests, examiners should merely be scorers, because "it will allow the scorer to observe and assess, free from the responsibility of keeping up the interaction with the candidate" (Harmer, 2014, p. 420).

In summary, the literature review unveiled numerous theories and hypotheses to explain SLA. Based on these, ELT methods thrived and transformed, from the grammar-translation method of old to more modern ones, such as CLT. No single theory or hypothesis is considered sufficient to explain SLA, nor is any single ELT method appropriate for fulfilling all learning objectives for all learners. However, the more recent ones are considered most effective. Despite its emphasis on teaching English holistically, the literature shows that CLT teaching and assessment of English speaking is still its Achilles' heel. Assessing oral communication is considered to be the "youngest subfield in language testing" (Fulcher, 2014, p. 13), and although it has steadily improved over time, reliable and authentic assessment of spoken language skills still warrant further research and attention.

Educational Assessment

Assessment

Assessment describes the collection and interpretation of evidence for making judgments or decisions, and guides teachers' instruction (Burke, 2010; Harlen, 2007). Its purpose is to determine how well students perform in terms of training skills and how much knowledge they've acquired from learning at a particular stage (Harmer, 2014; McNamara, 2000). Assessment can distinguish students' strengths and weaknesses and identify the gaps in their knowledge to guide instruction and interventions (Greenstein, 2012; Salend, 2009; Stigin & Chapuis, 2012). Different types of assessments can also increase student achievement and critically engage them (Mostafa, 2011). Ferrell (2012) stated that "assessment and feedback lies at the heart of the learning experience and forms a significant part of both academic and administrative workload. It remains, however, the single biggest source of student dissatisfaction with the higher education experience". For this reason, assessment procedures should be fair, valid and reliable (Greenstein, 2012).

In education, assessment is defined as teachers' multi-level judgments, including judgments about curriculum objectives, assessment tasks, grading criteria, task assessment, and recording of students' achievement (Allal, 2013). Student achievement is boosted by practising and receiving formative feedback through assessment

(Torrance, 2007), characterised by clarity in assessment procedures, processes and criteria. Appropriate assessment methods, proper assessment conditions and interpretation of student performances are also essential (Killen, 2005). However, assessment is a complex phenomenon (Orrell, 2005); it not only defines the educational outcome but also the way students learn. Based on Campbell (2008), the complexity of assessment is illustrated in Figure 2.3 – the highlighted areas indicate the aspects relevant to this research.

Killen (2005) described assessment as a multi-purpose activity. Athanasou (1997) identified three original purposes of assessment: selection, certification and classification. More recently, other purposes have been included, such as diagnosis, grading, progression, program evaluation, and instructional improvement (K. Cox, Imrie, & Miller, 2014; Harlen, 2007). Purpose is related to whether assessment is formative or summative (Harlen, 2007). Formative assessment provides information about the learning process and helps make decisions to spark learning progress, hence it is called *assessment for learning*. Summative assessment provides a summary of students' achievement over a period of time, hence it is known as *assessment of learning*.

Assessments are aimed at providing learners with quality feedback that will enable them to revise their performance to achieve higher standards (Carless, Salter, Yang, & Lam, 2011). It is considered a measure of students' potential and achievement, but also of teaching quality (K. Cox et al., 2014). Additionally, "the end goal of assessment is improved educational outcomes for students" (Salvia, Ysseldyke, & Witmer, 2012, p. 9). Carless et al. (2011) maintained that video and audio recording of students' oral performances facilitates reflection and feedback. These authors also believed that the use of technology can extend dialogue for feedback, promote open sharing and enable ideas to be revisited (Carless et al., 2011, p. 402).

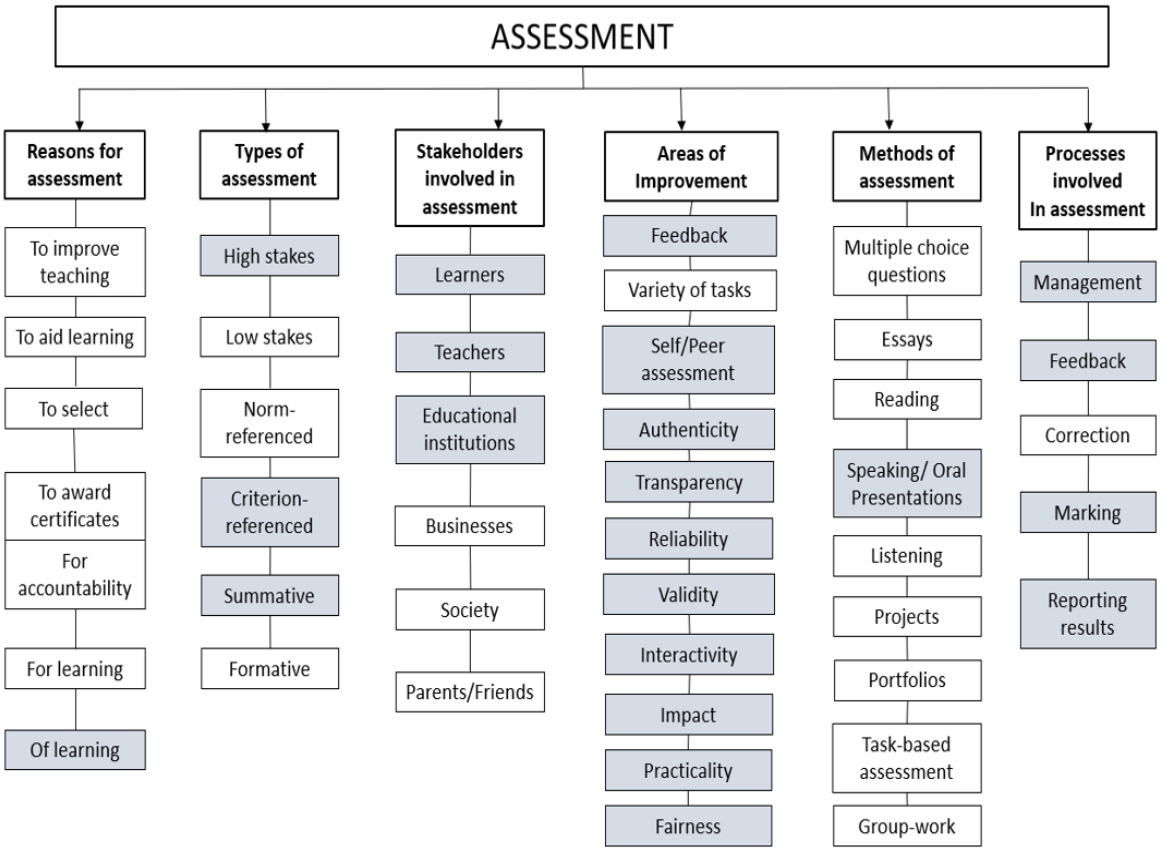


Figure 2.3 Complexity of Assessments.

Adapted from Campbell (2008).

Types of Assessment

Summative Assessment

Teachers use information derived from assessment to grade students before moving to the next, more advanced instructional unit. Administrators and policymakers use assessment scores to rank school achievement. Assessment that provides information about where students are at the end of the learning process is defined as summative assessment (Greenstein, 2010). Its purpose is to gather information on students' learning achievements, keep records of their learning progress, guide decisions for further study, and provide feedback and evidence of their progress to students and their parents (Harlen, 2007). The construct validity of summative assessment is higher than the construct validity of formative assessment, as criteria cover the full range of learning goals (Harlen, 2007).

Some scholars indicated that computer-assisted summative assessments generate considerable benefits, including automation, fairness and reliability in marking, prompt feedback, and flexibility in testing time and locations (Bernstein et al., 2010; Moere,

2010; Simin & Heidari, 2013). Learners are able to observe their progress during the assessment and their learning autonomy is encouraged (Kearney, Fletcher, & Bartlett, 2002; Simin & Heidari, 2013).

Formative Assessment

Summative assessment measures the product of students' learning i.e., what they have learnt; while formative assessment measures students' progress towards the learning goals i.e., how they learn. Formative assessment can inform students of their strengths and weaknesses and help them to improve their learning. Therefore, formative assessment is referred to as *assessment for learning* (Harmer, 2014).

Assessment Properties

Judging the effectiveness of assessment requires evaluation based on core criteria or properties (Harlen, 2007), such as validity, reliability, authenticity and accountability (Campbell, 2008; Miller, 2011). Reliability, validity and pedagogic impacts were the focus of this study and are discussed below.

Validity

Validity is an essential quality of assessment; it is understood that "a test is valid if it tests what it is supposed to test" (Harmer, 2014, p. 409). Validity relates to the decisions made from assessment information concerned with "whether the information being gathered is relevant to the decision that needs to be made" (Airasian & Russell, 2001, p. 16). That means validity of assessment refers to the appropriateness of the collected information, classified as highly valid, moderately valid, or invalid. There are four types of validity: construct validity, content validity, criterion validity, and face validity. A test which has criterion validity needs to produce similar results to other methods of measurement of the same abilities (Harmer, 2014).

Airasian and Russell (2001) highlighted three aspects of validity. First, whether assessment collects enough appropriate information for teachers to make the required decisions or not. Second, assessments that lack validity can lead to inappropriate decisions about learning and learners' achievements and may even be harmful. Third, all classroom assessment is concerned with validity, in particular summative assessment.

Reliability

Reliability "refers to the extent to which the results can be said to be of acceptable consistency or accuracy for a particular use" (Harlen, 2007, p. 21). The results of

assessment should be consistent, regardless of agencies or circumstances involved. The importance of reliability differs depending on the purpose of the assessment. Summative assessment requires higher levels of reliability than formative assessment.

Reliability of assessment is not concerned with the appropriateness of the information collected, but instead, relates to consistency, stability, and typicality of the information. Airasian and Russell (2001, p. 18) declared that "all assessment information contains some error or inconsistency; thus, validity and reliability are both a matter of degree and do not exist on an all-or-nothing basis". Reliability can be enhanced by providing clear instructions and ensuring consistency of the test conditions. It is also affected by the way tests are marked and the people who mark them (Harmer, 2014).

Pedagogic Impact

Assessment usually has an impact on curriculum and pedagogy because "what is assessed influences what is taught and how it is taught, and hence the opportunities for learning" (Harlen, 2007, p. 25). Assessment also has a powerful effect on what happens in classrooms, as "teaching and learning often reflect what the tests contain" (Harmer, 2014, p. 410). This reflection is called a washback or backwash effect. Figure 2.4 demonstrates the relationship between assessment, curriculum and pedagogy (learning and teaching).

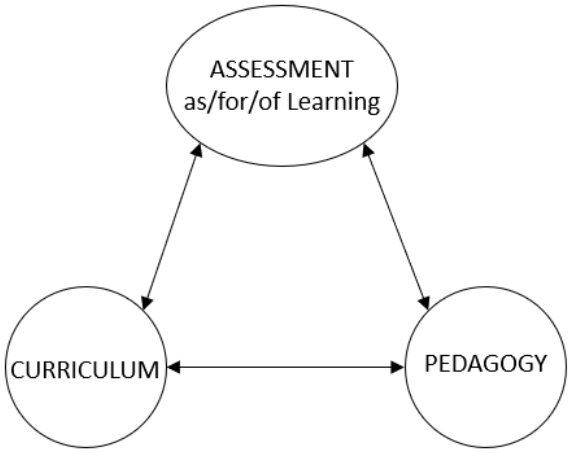


Figure 2.4 Relationship between Assessment, Curriculum and Pedagogy.

Based on Campbell (2008) and Harlen (2007).

The relationship between assessment and learning is complex and sometimes narrowly defined as *assessment of learning*, which mainly refers to marking and grading (Campbell, 2008). This definition has been expanded to include *assessment for learning* and *assessment as learning*. Either way, it is undeniable that assessment shapes the

learning process and is not separate from learning (Mikre, 2010). Evaluations during assessments are governed by the consequences of decisions that are made to students' individual learning (Fulcher & Davidson, 2007). While there is a plethora of literature on how to assess knowledge (Harlen, 2007; Heaton, 1990; McGaw, 2006; Reynolds et al., 2010), the literature on how to assess students' English speaking performance is more limited.

Theoretically, assessment and pedagogy follow the curriculum, in other words, methods of teaching and assessment are appropriate to what students are expected to learn (Harlen, 2007). Mikre (2010, p. 102) defined "assessment as a process for obtaining information on curriculum operation in order to make decisions about student learning, curriculum and programs, and on education policy matters". It therefore stands to reason that effective and reliable assessment will have a positive impact on both teaching and learning.

Performance Assessment

Performance assessment "involves students in activities that require them to demonstrate performance of certain skills or to create products that demonstrate mastery of certain standards of quality" (Stigin & Chapuis, 2012, p. 138). Grading performance assessment involves observation or examination of students' outputs. Students are asked to perform live and raters observe and make judgments. However, there is a risk of biased assessment due to the subjectivity of individual raters. Strict criteria should be established to enhance reliability of performance assessment.

More recently, performance assessment has received closer attention. One reason is that "unlike current tests that focus on facts and discrete skills, performance assessments are designed to test what we care about most – the ability of students to use their knowledge and skills in a variety of realistic situations and contexts" (Hart, 1994, p. 40). Performance assessment brings authenticity into the classroom by introducing real-world challenges and problems, and students often work collaboratively to find acceptable solutions. Performance assessment is believed to provide reliable information about student achievements that matches valued targets, including knowledge, performance skills, reasoning, and products (Stigin & Chapuis, 2012).

Second or Foreign Language Assessment

Second language assessment is defined as a process of gathering information about how much language a learner knows and can use (Isaacs, 2016). Language tests show

students their progress on the way to reaching fluency and proficiency. Tests can motivate students to achieve more, but also shows up their difficulties in acquiring a new language. Test results allow teachers to clearly see the problems and make in-time adjustments to their teaching and support of students (Fulcher & Davidson, 2013). It is also easier to group students based on test results and place them in suitable classes or levels (Chiedu & Omenogor, 2014; Crusan, 2012). Bachman and Palmer (1996) emphasised four major characteristics of language tests: construct validity, reliability, authenticity and interactivity. Chiedu and Omenogor (2014) added that besides validity and reliability, impact, practicality, transparency and fairness are also important qualities of language assessment.

According to Fulcher and Davidson (2007), there are three types of validity in language testing: criterion-oriented validity, content validity and construct validity. Criterion-oriented validity is the connection between the test and a common criterion, whereby the test score is compared to a criterion that measures the language competence of a learner, recognised on a larger scale beyond merely one organisation. Without criteria, judgment becomes subjective and unreliable. Content validity is the connection between the test and the target knowledge. Construct validity is the ability to accurately and consistently measure abstract ideas involved in tests, with "the quality of a test that allows us to make interpretations of the scores on the test" (Young & He, 1998, p. 2).

The reliability of assessment is reflected in consistent achievement in similar situations (McAlpine, 2002). Reliability is also an accurate measure of learners' competence, regardless of how the test is marked or who marks it. Factors that determine the reliability of language assessment include consistent scoring and the quality of test administration procedures (Chiedu & Omenogor, 2014). Moreover, the consistency of measurement determines the reliability of a language test (Bachman & Palmer, 2010). The consistency of measurement relates to the extent to which a test measures, and "a measure is considered reliable if a person's score on the same test given twice is similar" (Chiedu & Omenogor, 2014, p. 5).

Four different methods identify whether a language test is reliable or not (Chiedu & Omenogor, 2014): inter-rater reliability, parallel forms, item reliability and test-retest. This study adopted parallel forms as the research design and measure of test reliability. According to Chiedu and Omenogor (2014), the parallel form is "a measure of reliability obtained when a language teacher creates two forms of the same test by varying the items slightly. Reliability is stated as a correlation between scores of Test 1

and Test 2" (p. 6). Certain other factors, such as length of the assessment, clear instructions, fatigue, stress, motivation and environmental distractions can also affect reliability of language tests.

Authenticity is the degree of similarity between assessment tasks and real-life tasks in the target language (Frey, Schmitt, & Allen, 2012). Yujing Zheng and Iseni (2017) argued that authenticity in language testing should have an equal role to other factors, such as validity, reliability, interactivity and practicality. Interviewing to assess learners' speaking performance offers much authenticity, however, in such a context it is subjective and relative (Yujing Zheng & Iseni, 2017). Subjectivity lies in the way the test is designed and the way the test taker understands the test. Relativity refers to the way authenticity is perceived as more or less, rather than authentic or inauthentic (Bachman & Palmer, 1996). Yujing Zheng and Iseni (2017, p. 13) claimed that authenticity not only includes developing the test task and the test taker's interaction with the test task, but also scoring, by adopting authentic scoring criteria which are appropriate for judging fulfilment of real-world language use tasks.

According to Fulcher and Davidson (2007), interaction between teachers and students helps teachers to assess students' current abilities so that they can advise them what further learning should take place. Interaction demonstrates test takers' conversational strategies and provides evidence of their communicative competence. Interactivity not only describes the interaction between candidates and assessors, but also the knowledge of the test, language competence, performance strategies, and knowledge of the test topic (Bachman & Palmer, 1996; Young & He, 1998).

Another quality of language assessment is its impact on society, schools and stakeholders, including teachers and students. The decisions that are made based on test scores impact society, educational systems and individuals involved in the tests. Other factors, such as experience with taking tests and feedback also affect test takers (Bachman & Palmer, 1996). This is known as washback, defined as "the impact that a test has on the teaching and learning done in preparation for it" (Green, 2013, p. 40). Test design and how test takers perceive tests have an effect on their preparation. Teachers generally teach what is relevant to the test or "teach to the test" (Xie & Andrews, 2013), but Bachman and Palmer (1996, p. 33) recommended we "change the way we test" to ensure that assessment tasks are closely aligned with the instructional program (Bachman & Palmer, 1996, p. 33).

Practicality of language tests refers to their demand on resources as opposed to the availability of resources in the educational institution. These include human resources, material resources and time. Human resources are the test designers, invigilators, test scorers, and test administrators. Material resources are the test rooms, test materials and test equipment. Time resources refer to the available time for test development, implementation and scoring (Bachman & Palmer, 1996). Nicholson (2015) stated:

> Practicality refers to the economy of time, effort and money in testing and the consideration of resources is strongly linked to the financial costs involved in developing and administering a test. For a test to be practical it must be practical in terms of financial limitations, time constraints, ease of administration, scoring and interpretation (p. 223).

Fairness in language assessment is concerned with fairness to test takers (Kunnan, 2013). It stems from recognition of the fact that tests have the power to determine the future of an individual and may manifest as the inappropriate use of a test for different purposes (Shohamy, 2000). Shohamy (2000) suggested sharing the power among teachers and students by adopting multiple assessment processes, such as portfolios, self/peer-assessments, and observations to enhance test fairness. Above all, democratic and ethical assessment models in language assessment are vital for preventing misconstrued test results.

Computer-Assisted Language Assessment (CALA)

The use of technology in higher education and computer-based (CB) assessments are now commonplace in most university disciplines, including English (Newman, Couturier, & Scurry, 2010). For example, the TOEFL iBT tests have been delivered in 1,355 test centres in 149 countries. Pearson PTE Academic tests have delivered more than 27 million automatically scored test questions in CB test mode in over 100 countries around the world (Pearson, 2012).

Computer-Assisted Assessment (CAA)

Conventional paper-and-pencil assessments are time consuming and involve a significant amount of work to mark, deliver, and manage. Although paper-based tests are effective in some subjects for checking comprehension skills, they are not appropriate for evaluating performance. They are easy to grade, but this method only checks facts and memorised data and engages lower-level thinking skills, providing little evidence of what a language learner can actually do with the language (Rollings-

Carter, 2010). Things have changed from multiple choice and matching test designs to tests designed in digital formats and automatically graded, such as formal and informal online tests and quizzes (Gipps, 2005). Computers not only have the capacity to generate different versions of equally difficult tests, but also pose unique problems for students to practise. This method is known as computer-assisted assessment (CAA) or e-assessment (Ke, Yingwei, Xiaoli, & Yajun, 2011).

Computer-assisted assessment, sometimes referred to as computer-based assessment (CBA) or computer-supported assessment (CSA), is defined as the use of computers in assessing student learning (Bull & McKenna, 2004). Computer-assisted assessment is an alternative way of delivering paper-and-pencil tests. Since 1980, this digital testing method has changed significantly in regard to automatic evaluation, testing types, and integrated skills testing (Suvorov & Hegelheimer, 2014). With the integration of technology in teaching and learning, the potential to enhance intellectual capacity and creativity and prepare students to live in a technologically interconnected and globalised world (Chun, Kern, & Smith, 2016) has increased exponentially.

ICT-based assessment in higher education has developed from simple tasks (multiple choice, short responses) to various multi-media options, including audio and video recordings of student responses and productions as well as providing feedback (Gipps, 2005). There is also an increasing tendency to use ICT in test administration, because "results and statistics are immediately generated automatically and students obtain rapid feedback; exams can be easily stored and retrieved; and results may be further processed with other computer programs such as Excel and SPSS" (Mostafa, 2011, p. 3). Peer assessment and collaborative or group assessment via online chat-rooms, discussion boards and emails are all possible. The use of technologies in assessment is believed to enhance "the learning and teaching process and deliver efficiencies and quality improvements" (Ferrell, 2012, p. 3). However, automated marking of text and audio still has some way to go.

Gipps and Stobart (2003) agreed that feedback in the form of marks or grades alone does not enhance learning, while feedback in the form of comments encourages further learning. Some software products, such as TRIADS, QMark, and Online Assessment and Feedback, can provide automated feedback in online assessments, including diagnostic comments, showing the correct answers, and offering further explanation. Content-rich material and interactive web-based programs can be used to assess projects, case studies, essays, and group work, however, grading is done by hand in

these situations (Gipps, 2005). Automated scoring of complex responses remain challenging and need more research.

CAA covers different types of materials and reduces the burden on faculty and administrative staff, as well as offering flexibility (Ghilay & Ghilay, 2012) by transferring computerised tests to open access for students to use at home. Jamil, Topping, and Tariq (2012) concluded that some technological issues need consideration in order to realise the full benefits of CAA. For example, CAA requires investment in hardware, software setup and other facilities, yet despite some remaining limitations, CAA has increasingly been used in education to boost the efficiency of assessment (Abedi, 2014). Carr (2010) cautioned about the negative impact of technologies on student learning: "Our brains become conditioned only to accept and consume information in small, disjointed bits and eventually would not be able to process anything" (Carr, 2010, p. 130).

Growth of the internet and digital technologies has fuelled opportunities for online assessment methods. A large number of studies mentioned the benefits of online versus offline assessment, including improved student commitment, faster feedback (Baleni, 2015; Gikandi, Morrow, & Davis, 2011; Holmes, 2015), flexibility in place and time, and reduced marking time and administrative costs (Baleni, 2015). Hewson's (2012) study addressed concerns about the use of online course-based assessment methods and found that performance scores did not differ, regardless of whether the assessment was conducted online or offline. This quasi-experimental study supports the validity of online assessment by attesting to equal validity between online and offline assessment (Hewson, 2012).

Early research by Charman (1999) and Zakrzewski and Bull (1998) indicated that CAA generates significant benefits when used as a tool for summative tests, including automation, fairness and reliability in marking, prompt feedback, and the flexibility of testing time and locations. Kearney et al. (2002) confirmed that CAA provides learners with opportunities to study further and encourages student-centred learning. However, these researchers cautioned teachers against autonomous test generation from the same source, because it might encourage surface learning.

The advantages of using CAA in formative and summative assessments are widely believed to outnumber the disadvantages. In formative assessment, it allows for unsupervised study and enables learners to adjust their study in accordance with their

comprehension. In summative assessment, CAA allows learners to observe their progress during the assessment. This way of testing saves time on marking and reduces administrative work (Chalmers & McAusland, 2014).

Computer-Assisted Language Assessment (CALA)

Computer-assisted language assessment (CALA) is defined as a testing method that uses computer applications to elicit and evaluate learners' performance in a second or foreign language. Tools have been developed to facilitate the assessment of all language skills, including speaking and essay writing, but they have not been as successful in generating feedback on speaking tests and rating essays automatically (Suvorov & Hegelheimer, 2014). According to Winke and Isbell (2017), CALA is at the beginning of its development and language assessors are still attempting to incorporate technological advances into language testing.

Testing of vocabulary, grammar and reading has benefited from the early integration of ICT in assessment. According to Pathan (2012), the integration of technologies in scoring objective tests (Yes/No, multiple choice, matching, drag and drop, gap filling, and True/False) started in 1935 in the USA, with the use of the IBM model 805 for marking multiple choice questions. Winke and Fei (2008) stated that technologies enforce fast delivery and facilitate remote administration.

Online tests serve different purposes: replacement, proficiency, and selection for different levels. Web-based programs offer tests on reading, writing and speaking and a large collection of listening, reading, grammar and vocabulary tests. Pathan (2012) claimed that "the Web of many useful computer-adapted tests [CATs] and web-based tests [WBTs] are constantly growing and computers are used not only for test delivery but also for evaluation of complex types of test responses" (p. 33).

Pérez-Marín, Pascual-Nieto, and Rodríguez (2009) examined different computer-assisted assessment approaches to free-text answers for writing and speaking assessment, including short answers and essays. Despite criticism about assessing essays digitally, they found the development of natural language processing, e-learning, and the use of several automatic analysers, raters, and marking engines had rendered the idea feasible in practice. One example of positive change in the use of computers for essay scoring is the e-rater scoring engine, created by the Educational Testing Service (ETS) in the United States and used since 1999 to score GMAT and TOEFL. It is a powerful tool for evaluating essay-writing skills, capable of pinpointing grammar,

vocabulary, spelling and writing styles that need improvement. Based on natural language processing (NLP), this scoring mechanism increases scoring validity and reliability. However, Winke and Fei (2008) claimed that feedback generated by automated scoring engines is limited and argued that e-scoring should only be used for self-assessment.

In response to improving speaking assessment, Heaton (1990) suggested using a language laboratory to deliver speaking tests to a large number of students in a short period of time (five or ten minutes for each batch) instead of the usual time-consuming individual tests. He acknowledged that pre-recorded questions in speaking tests would never be as good as face-to-face interviews, because the scenario in which a student talks to a machine is not a natural, authentic situation. The inability to see the person talking and listening without a script, which means that the recorded questions keep going regardless of what the student has said, are said to be the limitations of this approach. However, audio recordings also offer a great deal of benefits; for example, a hint or prompt for the answer can be whispered, including asking the price, telling the time, and giving directions. Heaton (1990) argued that once all the drawbacks of this method were eliminated, it would be an effective way of delivering speaking tests.

In speaking assessments, "technology is seen not as a replacement for current methods, but as a new additional possibility" (Galaczi, 2010, p. 26). Despite the fact that no machine can replace a human, the development of technologies brings computer-assisted assessment closer to those conducted by humans. Improvements in speech recognition and natural language processing technologies have contributed to developments in oral language assessment and computerised speaking tests (Zhou, 2015).

Moere (2010) contended that computers are not capable of measuring social skills, such as nuances, politeness, turn-taking and negotiation in human speech, which are important parts of communication skills and convey meaning. Similarly, Bernstein et al. (2010) pointed out that computers fail to evaluate the strategic and complex content of spoken language in real life situations. Nor are computers capable of measuring complicated responses (Xiong, Evanini, Zechner, & Chen, 2013).

Witt (2012) expected that a number of features would gradually become available for individual or combined research to measure pronunciation and evaluate complex spoken language for a high degree of reliability in oral assessment. Williams and Newhouse

(2013) concluded that digital representation of student performances could provide authentic, reliable assessment of academic subjects, including second language speaking assessment.

Digital Representation

Digital representation is an information technology concept, defined as the process of digitising data and presenting it as a series of numerical values. Data digitisation involves putting information in a format that can be read by computers. It is used for different purposes, including newspapers on the internet, telephone systems, videos on DVD, and facsimiles. Digital representation has significant advantages in providing highly accurate, timely and accessible data and is fast replacing the ageing analogue methods (Mahmoud, Pirovano, & Larrieu, 2014). Parker and Dhanani (2012) stated that "digital representation has opened up all sorts of new usages of video" (p. 1). Digital representation has been studied in different fields, including palaeography for analysing medieval scripts (Ciula, 2005) and microstructure in 3D (Groeber & Jackson, 2014). However, it requires a large bandwidth on a transmission line and sufficient storage capacity.

Although audio recordings provide a record of oral transactions, many researchers have criticised their lack of visual aspects (Simpson & Tuson, 2003, p. 52). Context and other unrecorded factors, such as gestures, body postures, facial expressions, eye contact, etc. are all essential factors that facilitate comprehension of audio records. For this reason, video recordings may be regarded as more complete records of oral transactions.

Digital Representation in Assessment

The use of paper and pen to assess performances such as dance, presentations, and communication skills still seems inadequate. These types of performances would benefit from digital support because it "provides the ability to capture student knowledge and performance using a number of media (text, images, sound, and video) and this provides an improved and more authentic method compared with the current paper-and-pen method of assessment" (Pagram, 2013, p. 211).

Using digital representation in educational assessment has been a topic of interest for several researchers. For example, Stables and Kimbell (2007) captured students' innovative performance in their e-scape projects, initially using digital cameras to create a photographic portfolio of students designing a prototype, and then hand-held digital tools (PDAs - Personal Digital Assistants) to record their performance simultaneously

on a web space where it would be accessible to students, teachers and assessors. The authors reported that the digital representation provided students with evidence of their performance and clues for developing their prototypes, positive motivation and engagement.

Another example was the use of video recordings for assessing teacher competence by Admiraal, Hoeksma, Van De Kamp, and Van Duin (2011), confirming greater reliability and validity through enhanced fairness, meaningfulness and transparency. These researchers demonstrated that video recordings collect evidence of assessment in the form of rich information related to competence and the context in which the competence is presented (Admiraal et al., 2011). Others argued that video recordings promote in-depth discussion, critical reflection and self-reflection that bring about educational benefits (Borko, Jacobs, Eiteljorg, & Pittman, 2008; Rosaen, Lundeberg, Cooper, Fritzen, & Terpstra, 2008; Santagata, 2009).

Newhouse and Cooper (2013) established the possibility of using digital representation methods instead of face-to-face conventional methods to assess Italian speaking performance. They believed digital marking was as reliable and valid as the conventional method, with the added advantage of being faster and more convenient (Galaczi, 2010). Teachers in the Italian study stated that the video recordings of student performances led to fairer assessments and acknowledged the enabling role of digital technologies in students' critical reflection on their performance. The researchers concluded that digital forms of oral assessment were technically manageable and pedagogically feasible.

In summary, digital representations and their potential benefits to assessment have been widely explored in relation to providing evidence of performance (Stables & Kimbell, 2007), promoting peer feedback and discussion (Borko et al., 2008; Rosaen et al., 2008; Santagata, 2009), enhancing fairness (Galaczi, 2010), and being technically manageable and pedagogically feasible (Newhouse & Cooper, 2013). Although the advantages of digital representation in educational assessments are undeniable, they have only been studied in a limited number of subjects. Research across a larger variety of subjects would be useful to discover as yet unknown advantages and disadvantages.

Theoretical and Conceptual Frameworks

Theoretical Framework

The theoretical framework for this study was based on the literature review. Key terms, concepts and relationships are presented in Figure 2.5. The overall concept of the study was second language acquisition as this formed the main purpose of both teaching and assessment activities. Sociocultural theory and the output hypothesis underpinned the theoretical basis for developing second language communication skills and served as guidelines for selecting assessment tasks and discussing the pedagogical impacts of the assessment method investigated in the study.

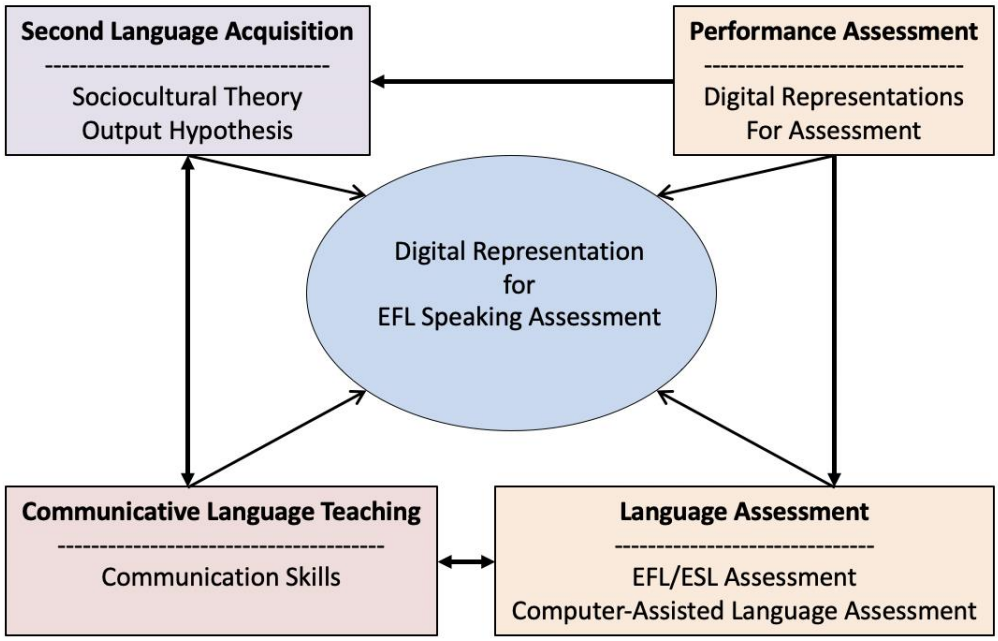


Figure 2.5 Theoretical Framework.

The literature review brought to light the dominance of CLT in second-language teaching for encouraging and improving learners' communication skills (Harmer, 2014; Jackman, 2016; Kayi, 2012; J. Richards & Rodgers, 2014). Hence, CLT served as the theoretical background for the selection of both assessment tasks and task assessments in this study, as well as providing guidelines for conducting authentic assessments.

The theoretical framework presents the relationship between *Performance Assessment* and *Language Assessment*. Assessing productive language skills, such as speaking and writing, is one type of performance assessment. Digital representations are frequently recommended in the literature for comprehensive and reliable assessment of performance (Borko et al., 2008; Galaczi, 2010; Newhouse & Cooper, 2013; Rosaen et

al., 2008; Santagata, 2009; Stables & Kimbell, 2007). Digital representation in second language assessment complies with and improves the quality of language assessment, bridges the gap between performance assessment and the assessment of EFL/ESL, and adds another choice to computer-assisted language assessment.

Technology Acceptance Model

The technology acceptance model or TAM (F. Davis et al., 1989) was adopted as a framework for this study (see Figure 2.6) to examine stakeholders' perceptions of computer-assisted EFL speaking assessment. TAM was commonly used in the field of psychology and originated from the theory of planned behaviour and the psychological theory of reasoned action (Marangunić & Granić, 2015). Today, it has become popular for exploring the behaviours of users in accepting or rejecting technology (Marangunić & Granić, 2015, p. 82).

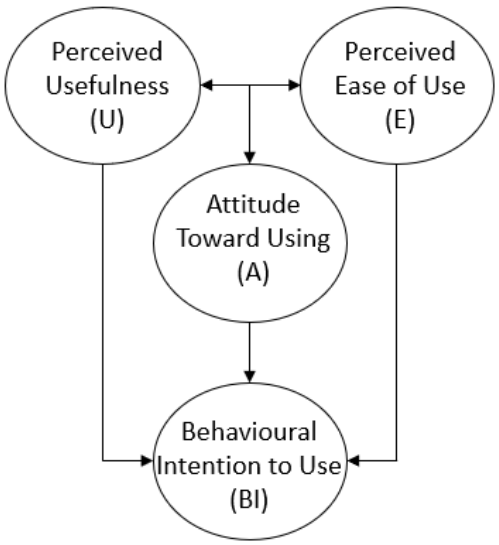


Figure 2.6 The Technology Acceptance Model.

Adapted from F. Davis et al. (1989).

TAM has evolved over three decades to include new factors; however, only four of the factors shown in Figure 2.6 were examined to align with the scope of this study. Perceived Usefulness (U) and Perceived Ease of Use (E) were singled out as two theoretical constructs that fundamentally determined the acceptance of using technology. U was defined as users' beliefs to the extent that the use of the technology would improve their performance (F. Davis, 1989; Pfeffer, 1982; Schein, 1980), whereas E referred to users' beliefs that the technology would be free from difficulties and effort (F. Davis et al., 1989).

As shown in Figure 2.6, U and E directly determined Attitude towards Use (A), where E was a determinant of U. The model indicates that all three factors (U, E and A) must be determined to identify Behavioural Intention to Use Technology (BI). BI was measured according to frequency of use, amount of time used, actual number of uses, and diversity of usage. U had a more direct influence on the emergence of BI (Lee, Kozar, & Larsen, 2003) – if users perceived the technology improved their performance, they had more intention to use it. E was found to be an antecedent of U and affected BI indirectly through U (F. Davis, Bagozzi, & Warshaw, 1992; Lee et al., 2003). In addition to these four core factors, other external variables affecting U, E, A and BI, such as stakeholders' technological literacy (Venkatesh, 2000), training (Igbaria & Iivari, 1995), computing support, experience (Chau, 1996), and availability of facilities (S. Taylor & Todd, 1995) were also investigated to better understand stakeholders' willingness and acceptance of digital assessment.

Feasibility Framework

The feasibility framework of Kimbell et al. (2007) was used in this study to inform the suitability of digital speaking assessment. This framework (see Table 2.3) was drawn from the findings of an e-scape project that examined e-solutions for creative assessments in a portfolio environment and extensive use of digital work in design and technology. The framework covers four key points: manageability, technology, functionality and pedagogy, as illustrated in Figure 2.7.

Table 2.3

The Feasibility Framework

Dimensions	Description
Manageability	Concerns issues of making such assessments do-able in normal classes, training implications for teachers and schools, and the scalability of the system for national implementation.
Technology	Concerns the extent to which existing technologies can be adapted for assessment purposes.
Functionality	Concerns the factors that an assessment system based on such technologies needs to address: The reliability and validity of assessments in this form, and the comparability of data from such e-assessments with non e-assessments.
Pedagogy	Concerns the extent to which the use of such assessment can support and enrich the learning experience.

It is popular in the field of performance assessment and e-assessment and was adopted as the principal guidelines for assessing technical systems construction in a 3-phase e-scape project in England (Kimbell, 2012a). It was also used to investigate the effectiveness of digital representations for assessing Applied Information Technology

(Newhouse, 2013), engineering studies (Williams, 2013), Italian studies (Cooper, 2013), and physical education studies (Penney & Jones, 2013). In these studies, manageability referred to the concept of making a digital form of assessment do-able in typical classrooms with a normal range of students. The other dimensions were unchanged from the original framework proposed by Kimbell et al. (2007).

The feasibility dimension of digital EFL speaking assessment is described in Figure 2.7. Manageability was analysed in terms of the do-ability of the assessment in normal classes, and the administration associated with assessment, including collection, storage and distribution of students' work and results.

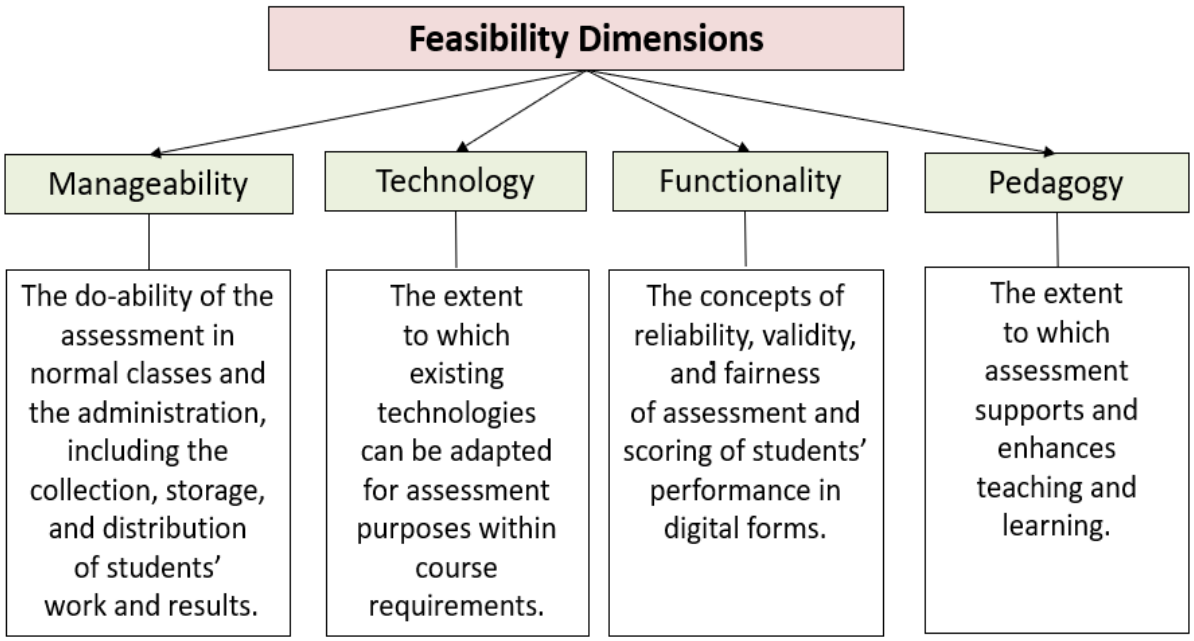


Figure 2.7 The Adapted Feasibility Framework.

The technology dimension covered the extent to which existing technological facilities and teachers' IT competence were compatible with the digital method for assessment purposes. Reliability, validity, and fairness characterised teacher and student perceptions of the functionality dimension and marking student performances in digital form. The extent to which assessment supported and enhanced teaching and learning was analysed as the pedagogic dimension of the study.

Research Framework

The literature review guided the research framework in Figure 2.8, depicting the key elements that formed the focus of the study and the relationships between them; i.e.,

using the digital representation method to assess EFL spoken language. The research framework indicates how the theoretical framework is utilized in the research.

As can be seen, the framework embodies the theory of second language acquisition, with the key concepts of sociocultural theory and the output hypothesis orienting the research. The assessment was conducted through the lens of communicative language teaching and principally targeted communication skills in an authentic teaching environment. The framework showed up the relationship between performance assessment and language assessment, with language assessment comprising one form of performance assessment.

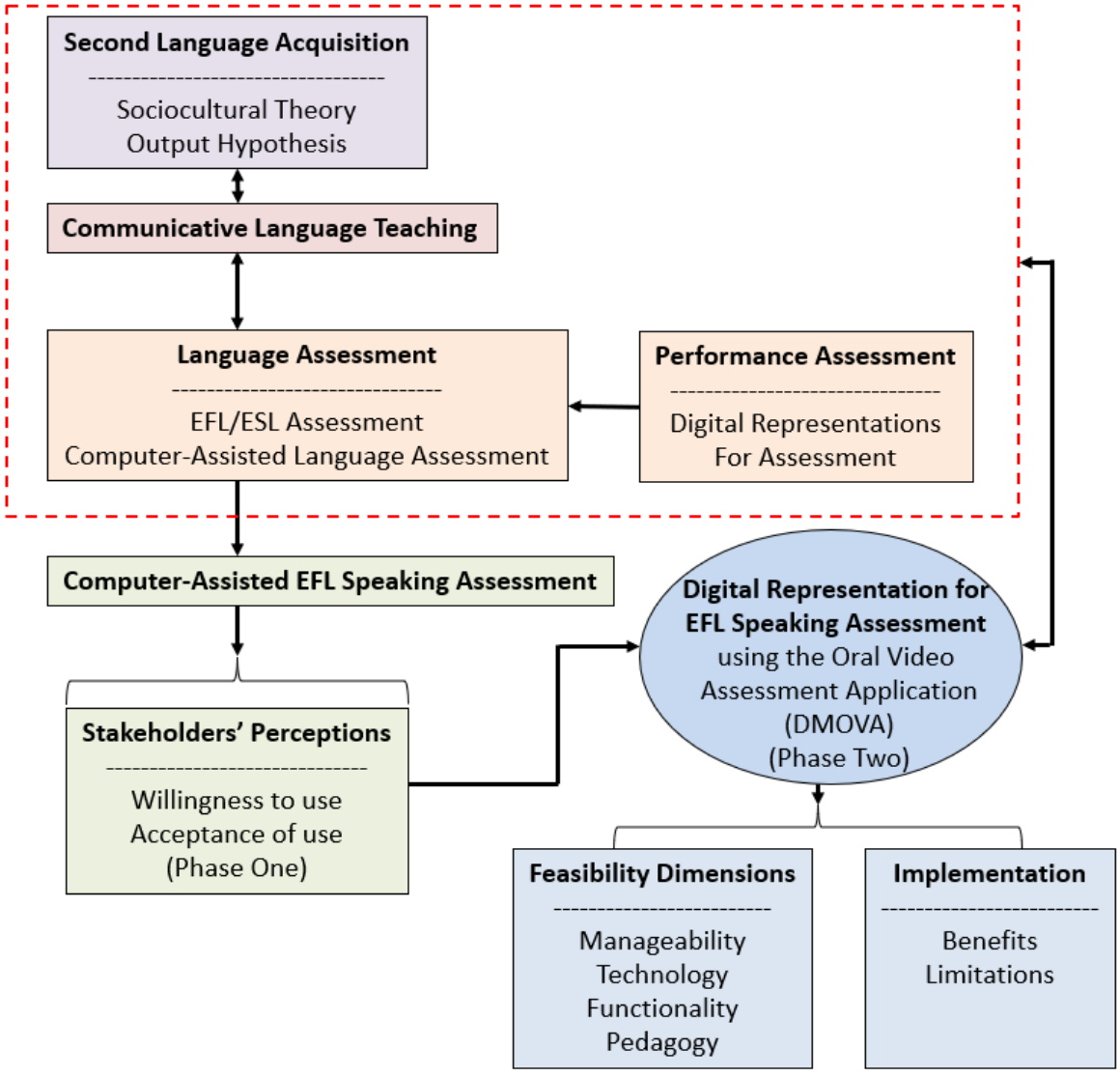


Figure 2.8 Research Framework.

The literature review indicated that computer-assisted language assessment was adopted as an alternative to paper-and-pencil language tests since 1935 (Pathan, 2012). Yet,

using computers to assess speaking has not gained the same popularity as for grammar and vocabulary, because of their inability to measure complicated responses and social skills (Moere, 2010; Xiong et al., 2013). Despite the limitations of computers for assessing speaking, it was nevertheless worthwhile to explore stakeholders' perceptions of computer-assisted EFL speaking assessment (Phase 1) to determine their willingness to use this method. The preliminary study led to the introduction of digital representation for EFL speaking assessment in Phase 2 using the Oral Video Assessment Application (DMOVA). A description of the Oral Video Assessment Application (OVA App) is provided in Chapter 3.

The feasibility of digital representation for EFL speaking assessment was analysed according to the four-dimensional framework of Kimbell et al. (2007), namely, manageability, technology, functionality and pedagogy. The benefits and limitations of implementation were also investigated. The findings of the study led to suggestions and recommendations for policies and practice of EFL speaking assessment using the digital assessment method.

Summary

The literature review covered two fields: English Education and Educational Assessment. Despite being an indispensable part of teaching, assessment is complex and diverse, and while teaching spoken English has received more and more attention, there is still no proper testing method that can measure this skill reliably. In addition, the exclusion of speaking proficiency assessment appears to be linked to the absence of an effective and scalable assessment method for enhancing reliability, fairness and authenticity, reducing administrative work, and saving resources.

The literature supports the idea of combining assessment with technologies to assess English speaking skills. While this is not a new concept, the most effective way of using technologies to assess speaking has yet to be found. The review also confirmed the potential for digital representation to enhance the reliability, transparency and fairness of assessments, provide evidence of performance and encourage reflection. However, further studies on the use of digital representation in EFL speaking assessment are necessary to draw verifiable conclusions.

CHAPTER 3
METHODOLOGY

The need to enhance Vietnamese students' English communication skills at all educational levels, particularly tertiary level, led the Vietnamese Ministry of Education and Training to introduce the National Foreign Languages Project 2020 (NFLP/ 2020 Project) in the Decision No. 1400/QD-TTg, titled "Teaching and Learning Foreign Languages in the National Education System, Period 2008 to 2020". Its purpose was to encourage English teaching and learning and achieve the goal outlined below:

> By 2020 most Vietnamese students graduating from secondary, vocational schools, colleges and universities will be able to use English confidently in their daily communication, their study and work in an integrated, multi-cultural and multi-lingual environment, making foreign languages a comparative advantage of development for Vietnamese people in the cause of industrialisation and modernisation for the country (MOET, 2008).

The project emphasised the task of renovating methods of assessment and grading in language training and proposed construction of an electronic databank to facilitate this goal. It called for teachers and assessors to actively apply Information Technology, not only in language training, but also in testing and assessment. The current research was conducted during enforcement of the National Foreign Languages Project 2020; its washback effect on the assessment of English language teaching and learning fully recognised by teachers, assessors and education administrators. In 2017, MOET assessed the NFLP/ 2020 Project and passed the Decision of Adjustment and Supplementation of the National Foreign Languages Project 2020 for the period 2017-2025 (MOET, 2017). The decision highlighted the need for improving assessment methods and integrating ICT into language assessment as one possible solution to improve language teaching and learning.

This study explored the potential of digital technologies to capture students' English speaking performances and more extensive use of digital assessment in English courses in Vietnam. It was partly motivated in response to the NFLP/ 2020 Project and the follow-up project of the Vietnamese MOET.

Theoretical Approach

This research project was conducted from a pragmatist perspective. According to pragmatic theory, researchers have the freedom to choose the methods, techniques and procedures most suitable for their research. Pragmatic researchers seek answers to "what" and "how" questions and use mixed methods to collect and analyse data, rather than one single approach such as qualitative or quantitative methods, because they believe that multiple sources of data will help them to better understand the research problem (Creswell, 2014b). Based on pragmatic theory, this study used mixed methods to collect and analyse the research data. Mixed methods are assumed to provide diverse types of data to foster a complete understanding of the research problem.

The research was conducted in two phases: Phase 1 was a survey that explored the perceptions of a particular population group and Phase 2 comprised interviews, observations, and intervention to further explore the impact of the phenomenon through case study analysis. The findings from Phase 1 informed Phase 2 of the study. The research design shown below was adapted from Creswell (2014b).

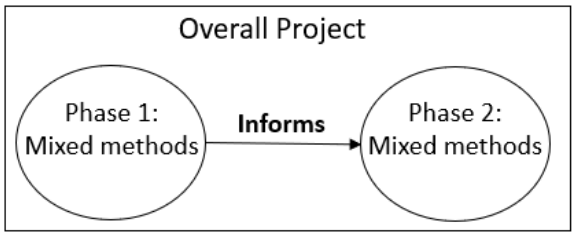


Figure 3.1 Two-Phase Mixed Methods.

Adapted from Creswell (2014b).

The overall objective of the study was to explore stakeholders' perceptions of computer-assisted EFL speaking assessment (Phase 1) to determine their willingness to use this method. The findings from Phase 1 informed the implementation of DMOVA (Phase 2). Both phases used mixed methods to analyse data, with each phase and method supporting and further explaining the other to create a whole picture and offer plausible answers to the research questions.

Mixed Methods

This research employed a mixed method design to collect and analyse data. Mixed method research is a combination of qualitative and quantitative approaches to provide a better understanding of the problem than can be provided by an individual approach

(Creswell, 2013, 2014a; Palinkas et al., 2015). Every method has its limitations; these can be mitigated by mixed methods to elicit more robust answers to research questions (Turner, Cardinal, & Burton, 2017).

A mixed method approach is not merely the collection of multiple forms of quantitative data from surveys and qualitative data from interviews or observation. It is the collection, analysis and integration of both qualitative and quantitative data sources (Creswell, 2014a). Thus, a mixed method design is not easy to implement, due to the amount of quantitative and qualitative data collected, and analysis that requires linking the qualitative and quantitative phases and integrating the results of both phases (Ivankova, Creswell, & Stick, 2006). The combination of qualitative and quantitative approaches in mixed methods improves the analytical power of the research (Sandelowski, 2000), since qualitative data support the analysis of quantitative data and vice versa (Clark & Creswell, 2008). For these reasons, mixed methods within a social science framework was appropriate for this study, supported by a congruent conceptual framework, data collection, analysis, and interpretation procedures (Creswell, 2013, 2014b).

Creswell (2009) proposed six basic mixed method designs. Concurrent triangulation was considered most effective for shaping the procedures of this study in relation to timing, weight, mixing, and theorising. It allowed the researcher to collect both quantitative and qualitative data simultaneously and reduce the time spent on data collection by not having to revisit the university. Two databases were analysed and compared to identify similarities, differences and combinations. In this way, the strengths of both qualitative and quantitative methods were harnessed to provide a comprehensive analysis of the research problem. The following figure illustrates the concurrent triangulation design.

According to Creswell (2009), concurrent triangulation offers flexibility and more options than other methods to analyse data in greater detail. It allowed the researcher to translate one type of data into another for merging, and then integrating and comparing the two databases side by side. Side-by-side integration entailed first introducing the quantitative results, followed by qualitative quotations to confirm or reject the quantitative results. In the current research, both data merging and side-by-side integration were used to interpret the findings.

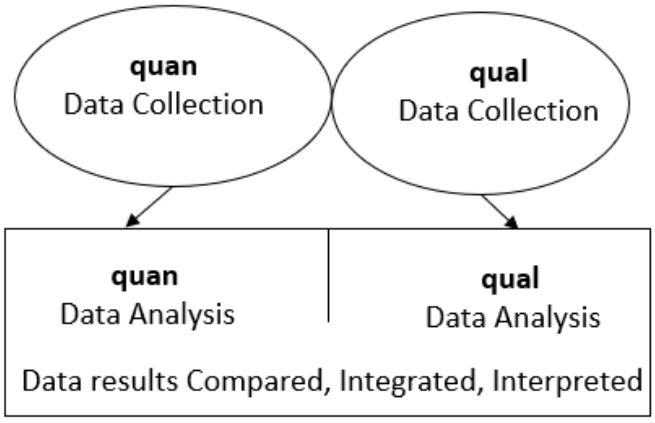


Figure 3.2 Concurrent Triangulation Design.

Adapted from Creswell (2009).

Numerous strategies ensured the validity of the data collected for this study, including audio recorded interviews, interview protocols; observations with video recordings; survey questionnaires with open and closed questions; multiple markers and peer markers, as well as triangulation of the data. The research used triangulation principles to optimise the mixed-method design and answer the research questions through better understanding and deeper insights (Burton & Obel, 2011). Triangulating the different methods used to examine the same research problem led to convergence of the data, increasing the credibility and reliability of the findings (Hesse-Biber, 2010). Figure 3.3 shows how triangulation works.

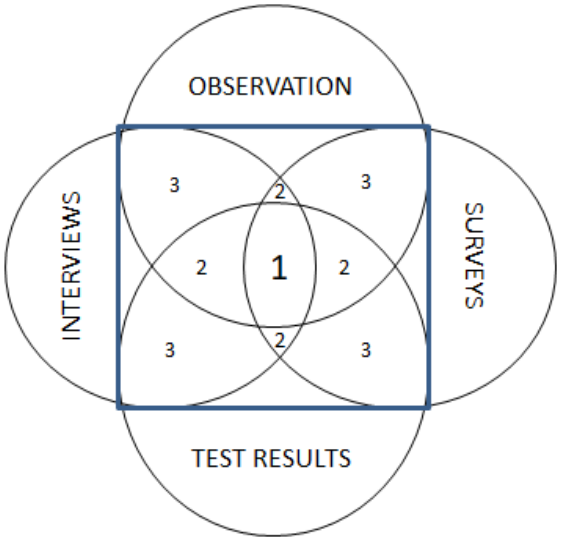


Figure 3.3 Convergence of Data Sources.

Data convergence occurs when similar findings show up in all or some of the different data sources. The current project collected data from surveys, interviews, observations

and the results of an English speaking test. The centre of Figure 3.3, marked 1, illustrates convergence of the findings after all the data were integrated. As can be seen, the findings from three data sources converged in the area marked 2, (Interviews-Observation-Surveys and Interviews-Surveys-Test Results), and from two data sources in the area marked 3. By interpreting these convergences, the results from the different data sources were integrated and validated. Convergence of the data sources is further discussed in Chapters 4 and 5.

Case Study

Case study design entails an intensive analysis and description of the research subject (Hancock & Algozzine, 2016). It can incorporate both qualitative and quantitative data collection methods and typically deals with a large amount of information. Case study is beneficial for describing real-life interventions, as it generates rich detail and depth of understanding (Yin, 2009). Given the nature of this research, case study methodology was an appropriate choice.

This project used descriptive case study to investigate the feasibility of digitising university students' English speaking performances for more reliable assessment. The focus was on summative, high-stakes, end-of-semester English speaking tests at university level. The test was high-stakes because the results determined whether students passed or failed English. The context or boundary of this case study (Hays, 2004) was an end-of-semester English speaking test undertaken by EFL students in three different classes and their teachers' marking practices. As the test takers, the students determined the case range, with teachers involved as English test invigilators and assessors of their live performances using digital representation. The participants of the case study possessed characteristics that could possibly be generalised to the whole population, i.e., university EFL teachers and students in Vietnam.

Sampling

The appropriateness and suitability of the sampling strategy (Cohen, Manion, & Morrison, 2011) is equally critical to the quality of a study as instrumentation and methodology. Cohen et al. (2011) recommended five key factors be taken into consideration:

- Sample size
- The representativeness and parameters of the sample

- Access to the sample
- The sampling strategy
- The kind of research method adopted: quantitative, qualitative or mixed.

Clearly, researchers cannot access the whole population because they are limited by expense, time, accessibility, the number of researchers and resources (Cohen et al., 2011). The sample size is also determined by the number of variables to be analysed. Cohen et al. proposed:

> There is no clear-cut answer, for the correct sample size depends on the purpose of the study, the nature of the population under scrutiny, the level of accuracy required, the anticipated response rate, the number of variables that are included in the research, and whether the research is quantitative or qualitative (Cohen et al., 2011, p. 144).

The most essential factor when recruiting a sample is that it should be representative of the whole population from which they are taken (Cohen et al., 2011). Samples can be recruited by means of probability or nonprobability sampling. Although nonprobability generates cost and time savings (Battaglia, 2008), it does not provide participants with equal opportunities to be included in the research. Purposive and convenience sampling are both nonprobability sampling techniques. Purposive sampling is sometimes criticised for being subjective and requiring expert judgment in its selection mechanism but is highly recommended for fostering deep understanding. Convenience sampling is also commended for the ease with which a sample can be acquired in terms of location, access and cost. Nonprobability sampling is popular with Web surveys where it is used as a form of snowball sampling because it reduces cost and time (Battaglia, 2008).

The benefits of purposive sampling are listed below. Based on the nature, purpose and research questions, it was selected for recruiting participants in the current study.

- It involves a wide range of participants with different experiences and perspectives related to the topic and therefore provides greater understanding of the subject;
- Selected participants can share similar ages, cultures, life experiences, traits and characteristics related to the research topic; and
- Participants can be chosen according to standard or typical characteristics within the population.

Convenience sampling offers both easy access and savings in terms of location and time (Etikan, Musa, & Alkassim, 2016). During the process of sample selection, representativeness of the larger population was taken into account to reduce bias, enhance the quality of the data, and increase the generalisation of the findings.

The target population, EFL teachers and students, was determined by the research questions and the nature of the study. All EFL teachers at FPT University were invited to participate in both phases of the research. To comply with the requirement of a large sample size for the survey in the first phase of the study (Cohen et al., 2011), participants were selected from the accessible population. Together with new participants, voluntary participants from Phase 1 made up the target population of the research. Phase 2 participants comprised students in three classes that were using Top Notch 2, Top Notch 3, and Summit 1 textbooks, equivalent to the three English levels: Pre-intermediate, Intermediate and High-Intermediate (see Appendix A). Table 3.1 shows the total number of research participants.

Table 3.1

Research Sample Size

Research Phases	**Teachers**	**Students**
Phase One	17	278
Phase Two	18	60

Instruments

Survey Questionnaire

Surveys are an effective method of collecting data about people's feelings, preferences, behaviours, and opinions on values (Fink, 2012). They offer flexibility and a straightforward way to collect data (De Vaus, 2013). In the form of online questionnaires, surveys are also suitable for research conducted in another country, hence, they were considered an appropriate data collection instrument for this study.

Survey questionnaires were utilised in both phases of the study. They were designed using Qualtrics, an online survey program, and contained both open and closed questions. Survey questionnaires are widely regarded as an effective tool for measuring participants' attitudes and eliciting other information anonymously. It is inexpensive, quick and easy for analysing closed questions, and provides "moderately high measurement validity for well-constructed and well-tested questionnaires" (Johnson &

Turner, 2003, p. 306). Online surveys offer electronic data entry, automatic data transformation into an analysable format, random question ordering, and other useful features to improve data quality and avoid errors (Van Gelder, Bretveld, & Roeleveld, 2010). However, response rates via email have proven to be unreliable (Groves, 2011; Hunter, 2012; Van Gelder et al., 2010), and there is also a risk of missing data, selective nonresponses, and vague answers to open questions.

To minimise potential weaknesses, the questionnaires were designed in accordance with the 13 principles of questionnaire construction proposed by Johnson and Christensen (2000). These were: questionnaire items matching the research objectives; understanding the research participants; using natural and familiar language; simple, clear and precise choices; avoiding loaded, double-barrelled and double-negative questions; mutually exclusive and exhaustive response categories for closed questions; multiple items for measuring abstract constructs; and pilot-testing the questionnaires.

The current study used a mixed questionnaire, defined as a self-reporting instrument, completed by the respondents (Johnson & Turner, 2003). It included open and closed questions, with one item text-enabled for further information and clarification by the respondents. There were Vietnamese and English language options for the surveys. Five Likert rating scales were incorporated to facilitate factor analysis. As recommended by Johnson and Turner (2003), the quantitative closed-question responses were supplemented by the rich, thick qualitative data gleaned from the in-depth interviews to best interpret the findings.

Semi-Structured Interviews

Previous studies on educational assessment used both questionnaires and semi-structured interviews to collect data (Brookhart & Durkin, 2003; Lai & Waltman, 2008). Interviews afford researchers the opportunity to probe participants for more detailed information that cannot be conveyed in questionnaires (Johnson & Turner, 2003). According to naturalism theory, interviews obtain deep meaning and help understand people's perspectives (Silverman, 2015) by generating rich data and enhancing data collection (McLafferty, 2004). Galletta (2013) recommended semi-structured interviews to allow room for participants to add new meaning to the research and for researchers to yield multidimensional streams of data. The author claimed that semi-structured interviews foster "a participant's responses for clarification, meaning making, and critical reflection" (Galletta, 2013, p. 24). Ensuring that semi-structured interviews yield

rich data, attention must be paid to preparation of the questions and development of the interview protocol.

In the current study, the semi-structured research questionnaire followed Galletta's (2013) guidelines. It included open questions probing participants' experiences related to digital performance assessment, specific questions to shed light on the complexities of the topic and concluding questions to help participants process and solidify their thoughts.

The semi-structured interview questions were posed in a way that encouraged engagement and meaningful responses. Interviews with teacher participants were intended to explore their experiences, attitudes, and recommendations regarding the digital testing method. The list of interview questions is provided in Appendix B.

Observations

Observation entails systematically gathering information specifically related to data obtained from surveys and interviews (Simpson & Tuson, 2003). "Observation is an important method because people do not always do what they say they do" (Johnson & Turner, 2003, p. 312). It offers the opportunity to collect additional valid and authentic data. Cohen et al. (2011) indicated that, in comparison to other research instruments, "the distinctive feature of observation as a research process is that it offers an investigator the opportunity to gather 'live' data from naturally occurring social situations" (p. 456), and researchers have opportunities to "look afresh at every behaviour that otherwise might be taken for granted" (p. 456) and "discover things that participants might not freely talk about in interview situations" (p. 456).

In this study, the observation instrument was set up to capture student and teacher behaviours and identify any technical issues during the EFL speaking tests. The tests were observed in actual, real time and video recorded, because video "offers a relatively 'unfiltered' record of all behaviours and transactions which occur in front of the camera, and a permanent, detailed record" (Simpson & Tuson, 2003, p. 51).

The observations were structured and focused on specific features of English speaking tests, including students' feelings of stress and confidence, and teachers' responses to the test procedures, test organisation and giving instructions. Other factors were also observed, such as technical issues, time taken for the actual test, and setting up for the test. All the categories were coded on observation sheets to facilitate observation, with the sheets designed to accommodate quick, freehand notes.

The categories for observing teachers were divided into four main themes:

1. Teacher behaviours towards operating the speaking test with a camera: This category was defined as teachers' positive and negative psychological behaviours in using the camera to capture student speaking performances, including displays of worry, stress, nervousness and confidence. Whether teachers had any problems with the presence of the camera was also explored.

   Teacher satisfaction and dissatisfaction with the digital testing method and their overall reactions were noted, as were expressions of pessimism and optimism about the testing method.
2. Test organisation: This referred to setting up for the test, including arranging the furniture in the test room, setting up the technologies, operating the camera to record student performances, and dividing students into groups for the group task. All evidence of ease and difficulty with conducting the tests was noted.
3. Teacher instructions: The rationale for observing teachers' instructions was to see whether it impacted on test results. The premise was that clear instructions led to better understanding by students and hence, higher test results, while on the other hand, the absence of clear instructions adversely affected student results.
4. Possible technical issues: The researcher observed no major technical issues, such as video recorder breakdowns, Wi-Fi interruptions, or software errors. Where technical issues did occur, the way they were resolved was noted, together with the outcome.

The categories for observing students were divided into three main themes:

1. Student behaviours in front of the camera and their attitudes toward the digital testing method: Just like the teachers, signs of positive and negative psychological behaviours by students were noted. Negative behaviours were characterised by worry, stress and nervousness, while positive behaviours included confidence, engagement in assessment tasks and cooperation. Any issues observed with students becoming accustomed to the presence of the camera were also noted in detail.

   Satisfaction and dissatisfaction were measured according to the student's ease and/or difficulty following teachers' instructions.

2. Student cooperation and engagement in assessment tasks: This aspect was related to students' attitudes. Positive attitudes were distinguished as the ease with which students engaged in discussion to demonstrate their proficiency and their cooperation in following teachers' instructions and rules. Difficulty getting involved in discussions and cooperating with one or more group members was identified as a negative attitude. Cases where one or two group members were dominant over others were also categorised as negative attitudes.
3. Time students started and finished the assessment tasks: Although time was pre-set for each assessment task in the OVA App, their starting and finishing times varied. The actual test time was calculated from when students started to speak until the time they completed the assessment task.

Previous studies showed that classroom observations can cause anxiety and stress for participants who may behave differently when they know they are being observed (Douglas, 1976; Jorgensen, 1989; Katz, 2015; Laurier, 2010). Consent letters (see Appendices C and D) were sent to potential participants with a clear and detailed explanation of how the classroom observation would be conducted. Teacher and student participants who were confident of behaving as usual in the classroom and willing to accept observations gave their consent.

The literature distinguished between overt and covert observations. In overt observations, participants know they are being observed, while in covert observations, participants do not know (Cohen et al., 2011). In this study, the observations were overt, i.e., the participants were aware they were being observed, according to the principles of informed consent and respect for their privacy and space. The unlikely potential for participants to experience adverse reactions was clearly explained, as were the benefits of the observations to the research. Participants were given time to consider before giving their consent.

The researcher was present and provided support during the test, assisting teachers and students to operate the technology, and on occasion, calling the next student into the test room. She was in the classroom 30 minutes before the test to familiarise teachers and students with her presence and helped set up the test room and the waiting room. Prior to the test, the researcher trained teachers how to use the camera recorder, and guided students to position themselves correctly in front of the camera for optimal visual and

sound recordings. During the training session, the researcher answered questions from both teachers and students, and communication was friendly and cooperative.

The researcher made her observations silently while sitting at the back of the classroom. Teacher and student behaviours were observed and recorded as codes on the observation sheets (see Appendices E and F). Other themes that were observed but uncoded were written down on the "further notes" section of the observation sheets. The video recordings were played and replayed after completion of the tests so that the researcher could record emerging codes and make additional notes. Analysing the observations entailed the researcher counting the frequency of references to individuals, groups, classes, events, activities, and behaviours and converting them into numbers (Cohen et al., 2011).

English Speaking Test

Tests are commonly used "to measure attitudes, personality, self-perceptions, aptitude, and performance of research participants" (Johnson & Turner, 2003, p. 310). In this research, tests were used to measure students' speaking performances via two different testing methods.

The test questions were derived from the Top Notch and Summit books published by Pearson Longman (see Appendices G, H, and I) and used to teach the students in this study. Prior to the tests, the class teachers reviewed and refined the test questions to ensure they were appropriate to what students were learning. The teachers returned a short list of questions to the researcher and these were used as assessment questions in the tests. The test questions were only revealed to students at the time of the test.

Students were grouped randomly from the name lists, resulting in a mixture of English competencies in each group. Four English teachers voluntarily acted as invigilators and agreed to observe and mark the students' tests.

Research Design

The study comprised two phases. Phase 1, the preliminary research, investigated teacher and student perceptions of computer-assisted speaking assessments. Their acceptance and willingness to use the new digital speaking assessment method was explored to inform Phase 2 of the study. Phase 2, the digitisation and assessment, was made up of two parts: first was video recording student performances for assessment and second

was teachers' marking of the recorded performances. The two phases are shown in Figure 3.4.

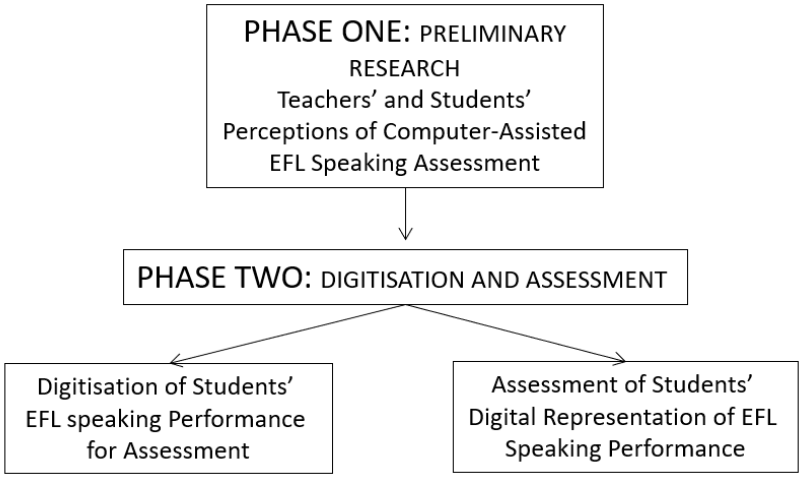


Figure 3.4 Research Design of the Study.

Phase One: Preliminary Research

Online surveys were used in Phase 1 to collect data about student and teacher perceptions of using ICT to support EFL speaking assessment. From this preliminary study, the researcher was able to measure their acceptance and willingness to experience an actual digital speaking performance assessment. Teacher and student survey questionnaires (see Appendices J and K) were designed using Qualtrics and delivered to participants online. They included closed and open questions to facilitate concurrent collection of qualitative and quantitative data. Data were collected and analysed in Phase 1 through a mixed method lens and informed the research in Phase 2.

Participants

An information letter was sent to all EFL teachers at FPT University explaining the survey and requesting they invite their class students to participate. The information letter doubled as an invitation to English teachers (22), of whom seventeen (17) agreed to participate and completed the online survey.

Phase 1 surveys were completed by 278 EFL students at FPT University, out of 365 invited. They were recruited by their English teachers who had forwarded on the information letter, in the form of an invitation, to their class students. Student participants came from IT Engineering and Business Administration majors. They were in their first year of university, attending an English preparation course before advancing to their major subjects in English.

Data Collection

The teacher survey contained twenty-two (22) questions (see Appendix J) and was estimated to take 10 to 15 minutes to complete. It contained closed questions, aimed at collecting demographic data on teachers' educational backgrounds; and open questions, for them to share their experiences, ideas, and initiatives. The data were analysed both quantitatively and qualitatively.

The student survey also contained twenty-two (22) questions and was delivered online (see Appendix K) using Qualtrics. Students were asked to share their experiences of using computers to take tests and their opinions of both paper-and-pencil and digital tests. On completion of the survey, they were asked to participate in the trial EFL speaking test using digital devices. The results are discussed in further detail in the introduction of DMOVA in Phase 2.

Data Analysis

In Phase 1 of the study, quantitative and qualitative data were collected. Numeric data derived from the closed questions in the survey were analysed quantitatively using descriptive statistics, while responses to the open questions were analysed using qualitative theme coding. Based on the technology acceptance model (see Figure 2.6) validated by (F. Davis et al., 1989), the core constructs for the themes of Perceived Usefulness (U) (see Table 3.2) and Perceived Ease of Use (E) (see Table 3.3) were used. Teachers' viewpoints on computer-assisted English speaking assessment were analysed using these constructs and examined in relation to their attitudes towards introducing DMOVA. Students' views about computer-assisted English speaking assessment were analysed using descriptive statistics and qualitative theme coding. Their attitudes towards the new testing technique were analysed and found to enfold a preference for computer-assisted English speaking assessment and conviction that digital testing was a viable option for this type of assessment.

Table 3.2

Constructs for Perceived Usefulness

Items	Perceived Usefulness
U1	Enhancing fairness
U2	Facilitating exam administration
U3	Improving the reliability of English speaking tests
U4	Offering authenticity
U5	Offering better interaction than face-to-face interviews
U6	Providing immediate feedback
U7	Reducing subjectivity in rating students
U8	Saving financial costs
U9	Saving time

Adapted from F. Davis et al. (1989)

Table 3.3

Constructs for Perceived Ease of Use

Items	Perceived Ease of Use
U1	Convenience in terms of test time and test locations
U2	Offering easy-to-use interfaces
U3	Providing recordings for later review
U4	Reducing stress and nervousness

Adapted from F. Davis et al. (1989)

Phase Two: Digitisation and Assessment

Participants

As shown in Figure 3.4, Phase 2 consisted of two parts. Part 1 involved digitising student EFL speaking performances for assessment by video recording their speaking tests. Part 2 entailed assessing the digital performances.

Sixty (60) EFL students from three classes/levels of English, namely, Pre-Intermediate, Intermediate and High-Intermediate, participated in Part 1 of Phase 2. All the students had agreed to participate in Phase 1 and Phase 2 of the study. They were joined by others who had consented to participating in Phase 2. Accordingly, not all the Phase 1 students participated in Phase 2, and not all the Phase 2 students participated in Phase 1.

Eighteen (18) EFL teachers at FPT University participated in Phase 2. They mainly comprised teachers who'd participated in Phase 1, supplemented by a newly recruited teacher. Four teachers, named T1, T2, T3 and T4, were voluntarily recruited to

invigilate, observe and live mark the tests in Part 1 of Phase 2. All 18 teachers were invited to contribute to Part 2 of Phase 2 as assessors of the students' digital performances. They all completed the survey, and 7 of them volunteered for a semi-structured interview with the researcher.

Part 1: Digitisation of Student Performances

This phase involved digitising the student speaking performances in a trial at FPT University, following the same procedures that were currently used by teachers and students, shown in Figure 3.5. The test included three activities: check-in to verify students' IDs, assessment task 1 (group discussion), and assessment task 2 (individual task). Student performances of the two assessment tasks were video recorded.

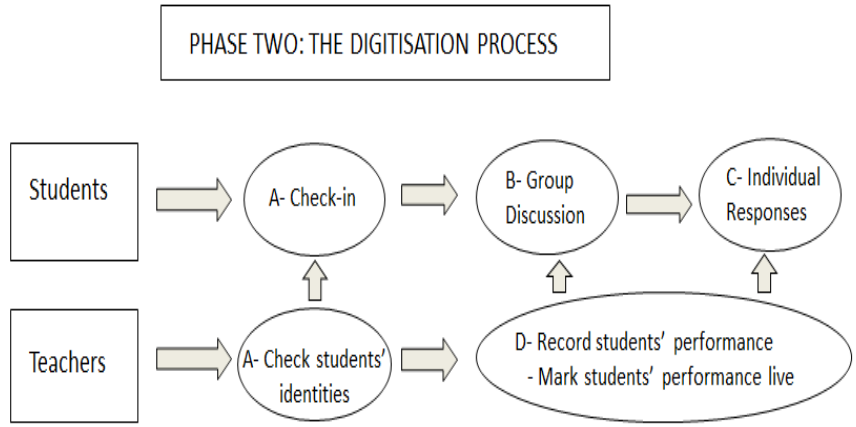


Figure 3.5 Phase 2 Research Design.

A - Student Check-In

Prior to commencing the speaking test, teachers checked students' names, photos, and ID numbers, and instructed them on the time they had for reading the test guidelines, preparing for and completing each task. Students were informed that they'd be reminded of time remaining and when time ran out for each task. Student check-in took approximately two minutes for each group of four students.

B - Group Assessment Task (6 minutes - plus preparation time of 4 minutes)

Students were randomly divided into groups of four from the student list. Each class included five to six groups, for a total of 16 groups altogether. Each group randomly chose a topic for discussion from a list of topics. After four minutes of preparation time, they discussed their chosen topic for a maximum of six minutes. Preparation time was necessary to appoint a group leader, decide the format of the discussion and organise their arguments. Their roles as group leaders did not add marks to their assessment

results. Students' English speaking competence was assessed according to the marking key in Appendix L.

C - Individual Assessment Task (3 minutes - no preparation time)

After completing the group discussion, each student undertook an individual assessment task by selecting a random topic and talking for a maximum of three minutes. Students were not permitted time to prepare, because the exercise was aimed at evaluating their instant responses to authentic communication situations. Figure 3.6 shows the position of the camera and the layout of the test room for the individual assessment tasks.

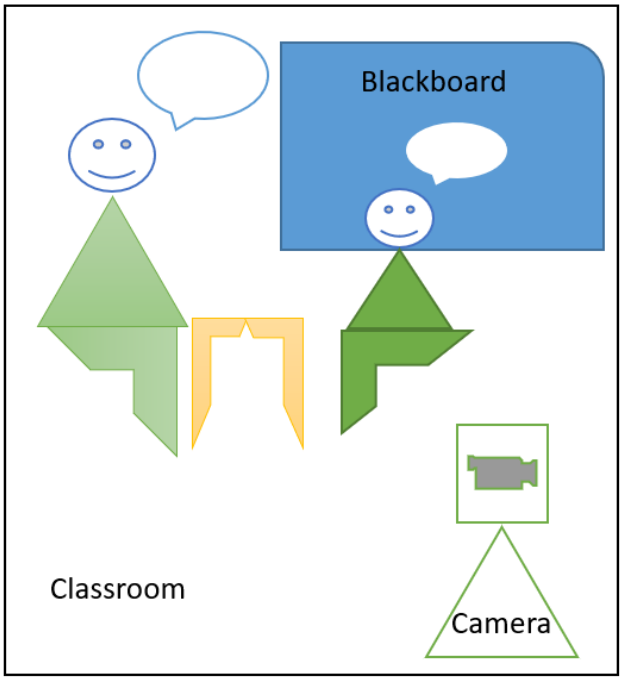


Figure 3.6 Layout of the Test Room.

D - Teacher Recording and Marking Activities

The schedule for the speaking tests was discussed with the teachers and implemented as shown in Table 3.4. As can be seen, two teachers invigilated each English speaking test. They were asked to record the student performances and mark then in the same way they usually marked speaking tests. Teachers were provided with a printed marking key (see Appendix L) and marking paper sheets (see Appendix M) for the two assessment tasks.

Table 3.4

Schedule of EFL Speaking Tests

Sessions	Class	Number of students	Invigilators
1	Intermediate	23	T1, T4
2	Pre-Intermediate	17	T1, T3
3	High-Intermediate	20	T1, T2

Part 2: Digital Assessment of Student Performances

The assessment phase involved all 18 teachers marking the video recorded student performances. There were 76 videos in total. Teachers T1, T2, T3 and T4 were each provided with an iPad to do their marking, and their test results were extracted from the OVA App. The other teachers were provided with an internet link, and a unique user name and password allowing authorised access to the digitised performance files in the Cloud. There were 16 recordings of group tasks and 60 recordings of individual tasks. Table 3.5 shows the teacher distribution for marking the digital performances.

Table 3.5

Teacher Distribution for Marking the Digital EFL Performances

Class	Number of students	Number of recordings		Teachers
		Group	Individual	
Intermediate	23	6	23	T1, T2, T3, T4, + others
Pre-Intermediate	17	5	17	T1, T2, T3, T4, + others
High-Intermediate	20	5	20	T1, T2, T3, T4, + others

Data Collection

Part 1: Observations and EFL Speaking Tests

In Part 1 of Phase 2, a speaking test was organised for three classes of 60 students and four teachers. The tests were conducted in the same way as they usually were at FPT University – students completed two assessment tasks while teachers observed and then marked their tests using paper and pencils. The entire process was video recorded. The presence of the researcher in the room was announced to both teachers and students before the test. During the test, the researcher provided technical support when needed, but otherwise sat silently in the far corner of the room without interfering. Observation data were noted on the structured observation sheets (see Appendices E and F).

Two teachers in each class marked the student performances in the usual way with paper and pencils. The test results were collected and transferred to an Excel

spreadsheet for data analysis. Figure 3.7 summarises the data collection process in Phase 2 of the study.

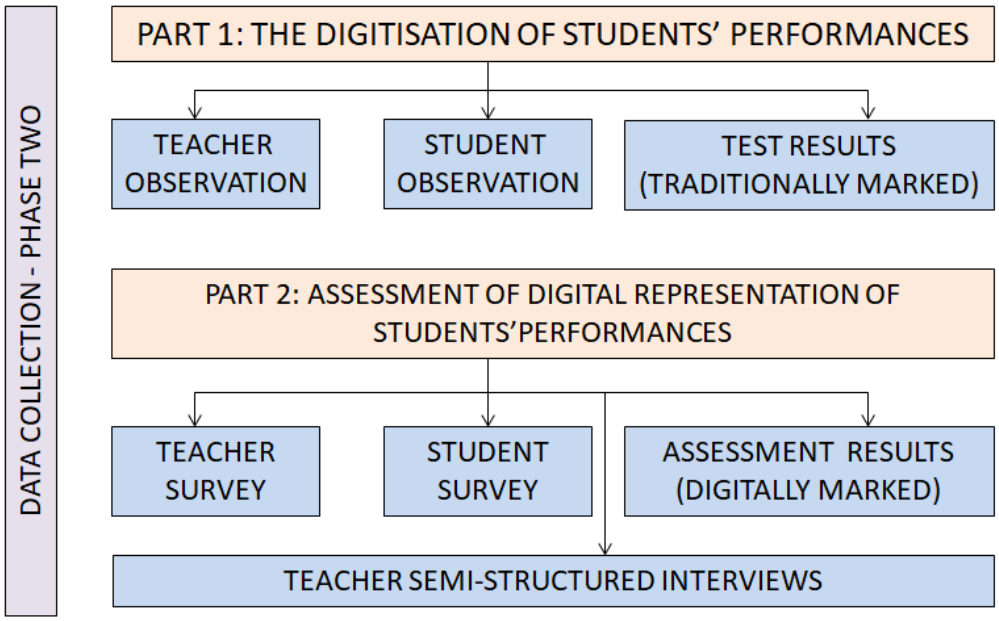


Figure 3.7 Data Collection Scheme in Phase 2.

Part 2: Surveys, Semi-Structured Interviews and Assessment Results

Eighteen teachers participated in Part 2 as assessors of student digital performances and marked on iPads. The results awarded by four teachers (T1, T2, T3, and T4) were recorded for correlation analysis. After they'd finished marking, the teachers were asked to complete a survey questionnaire (see Appendix N) and participate in semi-structured interviews with the researcher. Seven teachers agreed to be interviewed.

The video recordings were shown to the students so they could see their digital performance and understand the marking and feedback. They were then asked to complete an anonymous survey questionnaire (see Appendix O) delivered online to their email addresses.

Data Analysis

The data were analysed using mixed methods. Closed question responses in the surveys were analysed using quantitative statistical analysis. Open question responses from the surveys, the observational data, and semi-structured teacher interviews were coded qualitatively according to themes. NVivo and SPSS data analysis tools were used to interpret qualitative and quantitative sources of data. SPSS was also used to analyse correlations between the live and digital marking results. Data types and sources were

triangulated to enhance the credibility of the research findings. Figure 3.8 shows how the analysis of different data sources addressed the research questions.

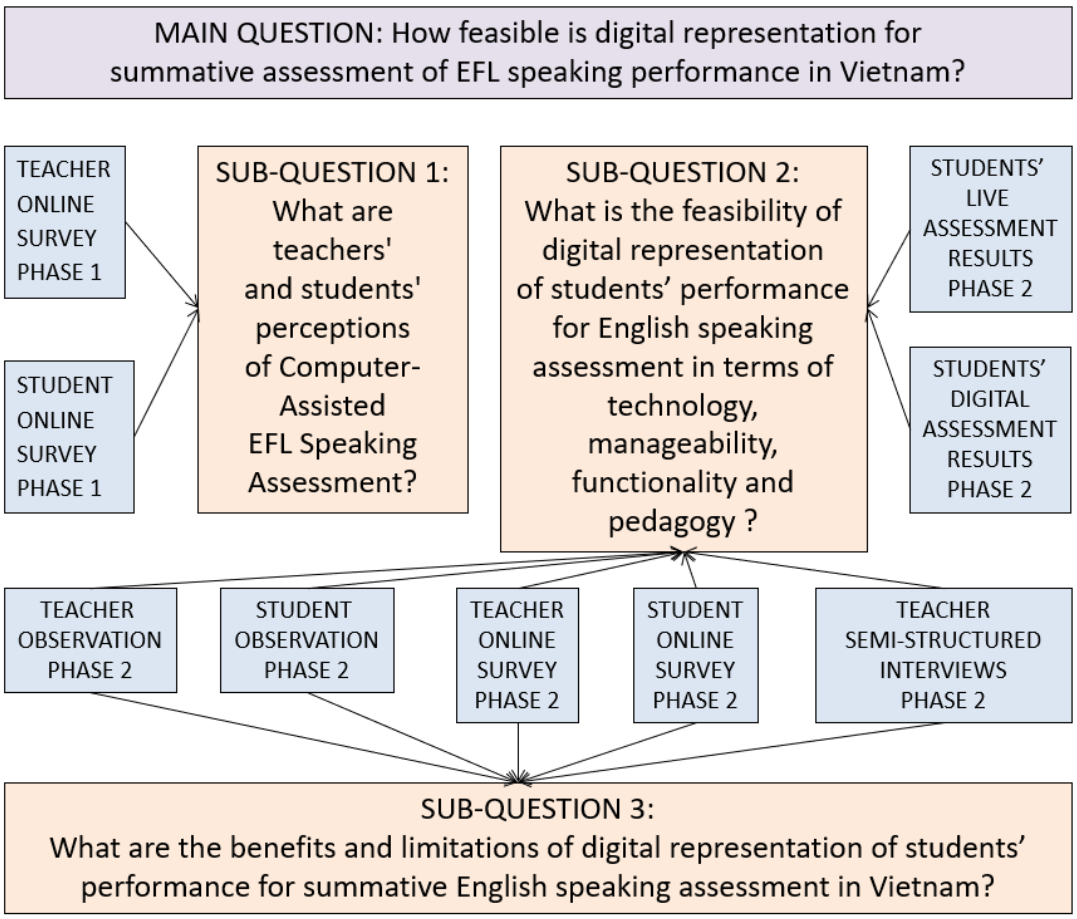


Figure 3.8 Data Sources for Answering the Research Questions.

The study made use of correlation tables to demonstrate consistency and similarities in the two methods of marking. They showed mean scores, maximum and minimum scores, and correlation coefficients, as well as highlighting similarities and differences between the marking results. This assisted in identifying significant discrepancies in the results awarded by the different teachers and differences in their personal judgments and standards in assessing English speaking skills.

Feasibility Analysis Framework

The qualitative and quantitative data collected from the observations, surveys, interviews and student assessment results were synthesised and analysed using mixed methods. Feasibility of the digital assessment method was measured according to a feasibility framework adapted from Kimbell et al. (2007), depicted in Figure 2.7.

As previously mentioned, the feasibility analysis framework measured the four different dimensions of manageability, technology, functionality and pedagogy. Manageability analysed the administration of assessments, including collection, storage and distribution of student work and results. The technology dimension assessed the extent to which current technological facilities and teachers' IT competence could be adapted to the digital assessment method. In the functional dimension, teachers' and students' perceptions of assessment reliability, validity and fairness were examined, as well as digital scoring of the student performances. The pedagogic dimension described the extent to which assessment supported and enhanced teaching and learning.

Cronbach's Alpha Reliability Coefficient

The survey questionnaires used a 5-scale Likert response system and multiple items rather than individual ones to increase reliability and validity (see Appendices N and O), as recommended by McIver and Carmines (1981):

> The most fundamental problem with single item measures is not merely that they tend to be less valid, less accurate, and less reliable than their multi-item equivalents. It is rather, that the social scientist rarely has sufficient information to estimate their measurement properties. Thus, their degree of validity, accuracy, and reliability is often unknowable. (p. 15)

A multiple item scale was developed for the teacher and student survey questionnaires to deeply explore participants' attitudes toward the existing and digital assessment methods. The multi-item questionnaire was purposefully designed to facilitate calculation of Cronbach's alpha internal consistency. Cronbach's alpha index was used to check the reliability of the variables to ensure consistency in the survey responses. Cronbach's alpha reliability coefficient ranges from 0 to 1, with high values indicating higher internal consistency of the items on the scale (Gliem & Gliem, 2003). The alpha values, based on George's (2011) alpha value table, are shown in Appendix P.

NVivo Theme Coding

Responses to the open questions in the survey, observational data and the teachers' semi-structured interviews were coded by emerging themes using NVivo 12.1.0, developed by QSR International. NVivo qualitative software was selected because it is a powerful coding tool capable of addressing threats to validity (Siccama & Penna, 2008), interrogating interpretations, scoping data, establishing saturation and maintaining audit

and log trails to ensure the data are used appropriately, the inquiry is thorough and leads to the best outcomes (L. Richards, 2004).

In this study, qualitative data were imported into NVivo as audio recordings, Pdf and Word files. Both independent and tree nodes were evident; the latter assisted with organisation, analysis, and modification of the codes throughout the study (Gibbs, 2002). The tree nodes were arranged in a hierarchical structure to indicate the relationships between the main themes and subthemes, moving from a general category (parent nodes) to a more specific category (child nodes). As proposed by Miller, Huberman, Huberman, and Huberman (1994), a variable-oriented strategy was used to search for themes across the files. This facilitated exploration of the data for specific perspectives, attitudes, reactions, similarities and differences, as well as relationships between parent and child nodes and connections between categories (Gibbs, 2002).

Audit and log trails were used to ensure consistency in the data collection and findings (Siccama & Penna, 2008) by "providing a means for tracking decisions and assumptions. It also allows outsiders to see how such decisions and assumptions have evolved over the life of the project" (Siccama & Penna, 2008, p. 100). In the current study, the audit trail included time and date stamps on documents before importing them into NVivo. Dates and times when databases were accessed and modifications made to the theme coding were also recorded and saved.

Descriptive Statistics and Correlation Analysis

SPSS was used in this study to generate bivariate correlations and descriptive statistics of the test results. Correlation is defined as a statistical way of looking at relationships; when two things are correlated, they vary together in the same direction (Schmuller, 2013). Correlation analysis has been widely used in the fields of language learning and teaching to investigate relationships between enhancement of learner autonomy and higher proficiency in the target language, e.g., Shukla (2018). The topic frequently appears in the literature on testing second language speaking (Fulcher, 2014).

A major challenge of this research was establishing the degree of agreement between results derived from existing and digital methods of assessing student performances. A correlation analysis helped to investigate the degrees of agreement and drew attention to correlations between marks awarded by multiple teachers using the digital marking method. The analysis also made it possible to determine the reliability of digital marking versus the existing marking method.

The purpose of correlation analysis is to support the validity of a particular hypothesis. The "validity argument for indirect speaking tests has been that they measure the same construct as direct speaking tests … The argument is that if scores on two tests are so highly associated that one can predict from one to the other, the test must be "construct-equivalent" (Fulcher, 2014, p. 172). The same author argued that more information is needed than just the number from +1 to -1 (Fulcher) to interpret a correlation coefficient. In this study, the correlation coefficients and validity of the correlation findings were confirmed and supported by triangulation with other data sources and adoption of different data analysis methods. Details are presented in Chapter 5.

Oral Video Assessment Application (OVA App)

Answering the research questions required a mobile application, developed in collaboration with the Centre for Schooling and Learning Technologies (CSaLT) at the School of Education, Edith Cowan University. CSaLT had carried out research in performance assessment and developed mobile performance applications to facilitate all areas of assessment. A customised mobile performance assessment application, named Oral Video Assessment Application (OVA App), was developed for this research to address the research questions in relation to its manageability, technology and functional dimensions. The OVA App was developed on FileMaker by Dr Alistair Campbell, from CSaLT, who was also a supervisor, program developer and application administrator for this research project.

Since the research focused on performance assessment of English speaking skills and was conducted in a particular research context, the OVA App needed to:

- Record student live English speaking performances in the real context of a test room,
- Facilitate the marking process and allow multiple markings of each performance,
- Provide easy access to the recordings for markers and reviewers,
- Enable easy retrieval and distribution of test results,
- Be compatible with the existing technological facilities and conditions at the university,
- Be user-friendly and suitable for teachers with low-level ICT backgrounds.

The OVA App was designed as a prototype and customised for the purposes and particular context of the research. Its features included videoing, marking, storing,

uploading, sharing, and exporting results to Excel. The OVA App operated in three environments: (a) on an iPad using FileMaker GO; (b) in a Windows or Mac environment using FileMaker software; and (c) in a browser. As a platform for collecting video data on student speaking performances with an embedded marking key, the App forged a new way of marking and providing feedback. Instead of using paper and pens, teachers could mark digitally at a time and place of their choosing. The App had three main functions: recording, marking, and managing – these functions are shown in Figure 3.9.

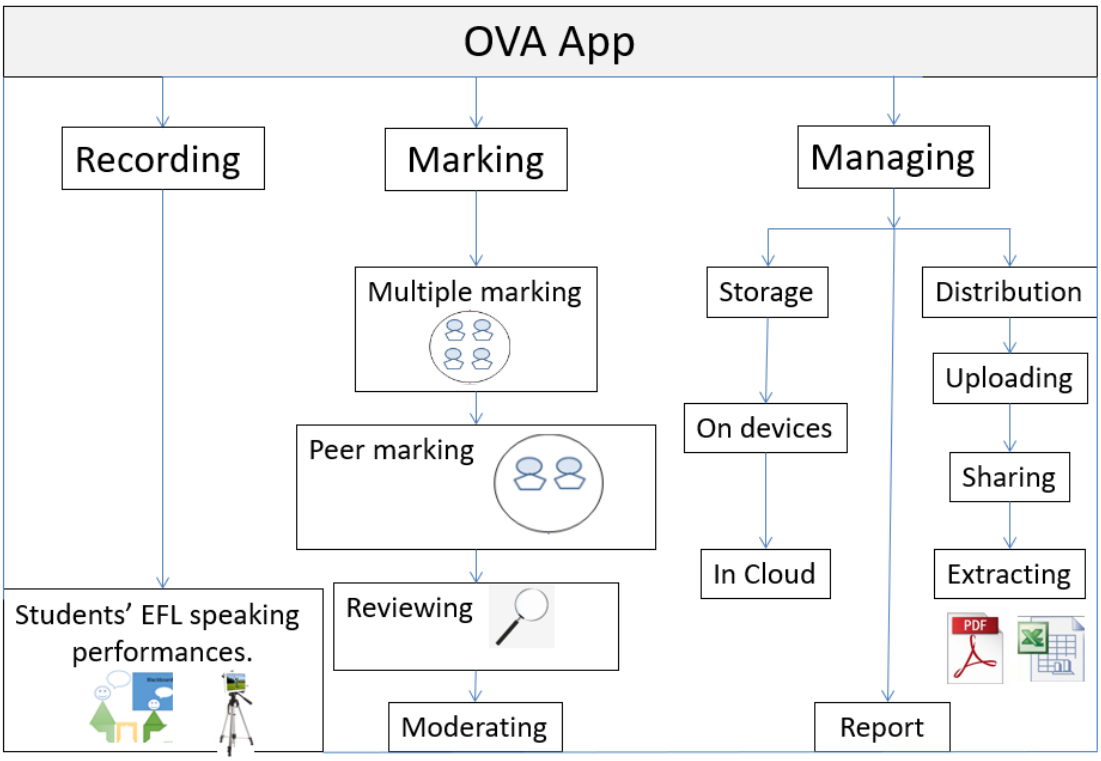


Figure 3.9 Main Functions of the OVA App.

The functions were displayed on the home page of the application (see Figure 3.10) and activated by different buttons, where other information provided an overview, brief explanation of the application's features and their purpose, as well as ethical information.

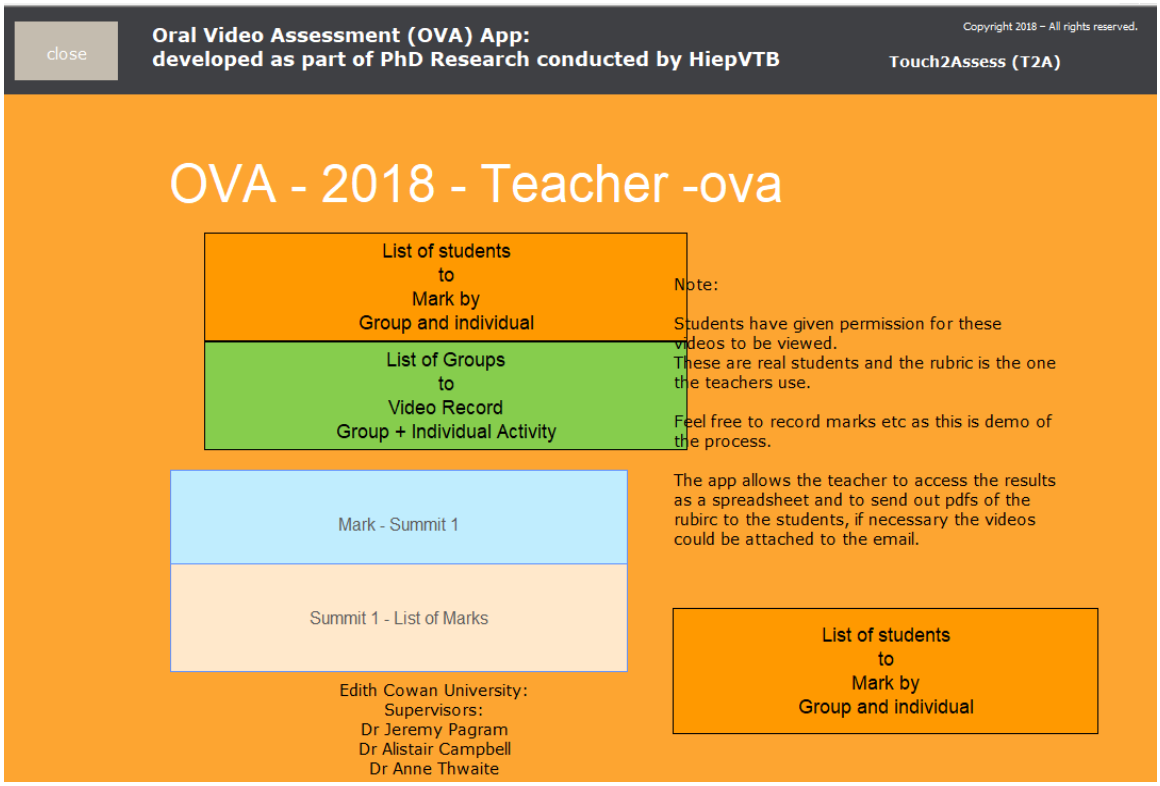


Figure 3.10 The Home Page of the OVA App.

As shown in Figure 3.10, teachers clicked on the green button, *Video Record Group and Individual Activity*, to open the video recording page and start recording. To mark students' performance, they clicked on the orange button, *Mark Group and Individual Activity*, which linked them with the database of video recordings. To check student results, teachers clicked on the white button, *Students' Results*, where they were displayed on spreadsheets with options to show results for separate criteria or total results. These functions are further described below.

Recording Function

The equipment needed to video record student speaking performances comprised an iPad with the OVA App installed and a tripod. Figure 3.6 shows the process of recording. The iPad was mounted on a tripod for video recording, and teachers simply opened the App on the iPad and pressed the start button. The height of the tripod was adjustable to cater for optimal visuals and good quality videos. While the App recorded, teachers took notes, asked questions and marked in the conventional way. The recording stopped automatically when the time was up for each assessment task, and teachers were able to manually stop the recording if students didn't reach their time limit.

As mentioned above, the green button, *Video Record Group and Individual Activity*, was linked to a page where teachers could access the videos of student performances.

The Video Recording function had an offline option that enabled recording of student performances without internet connection. Figure 3.11 shows the Video Recording Interface of the application with different colour buttons for different functions of the App.

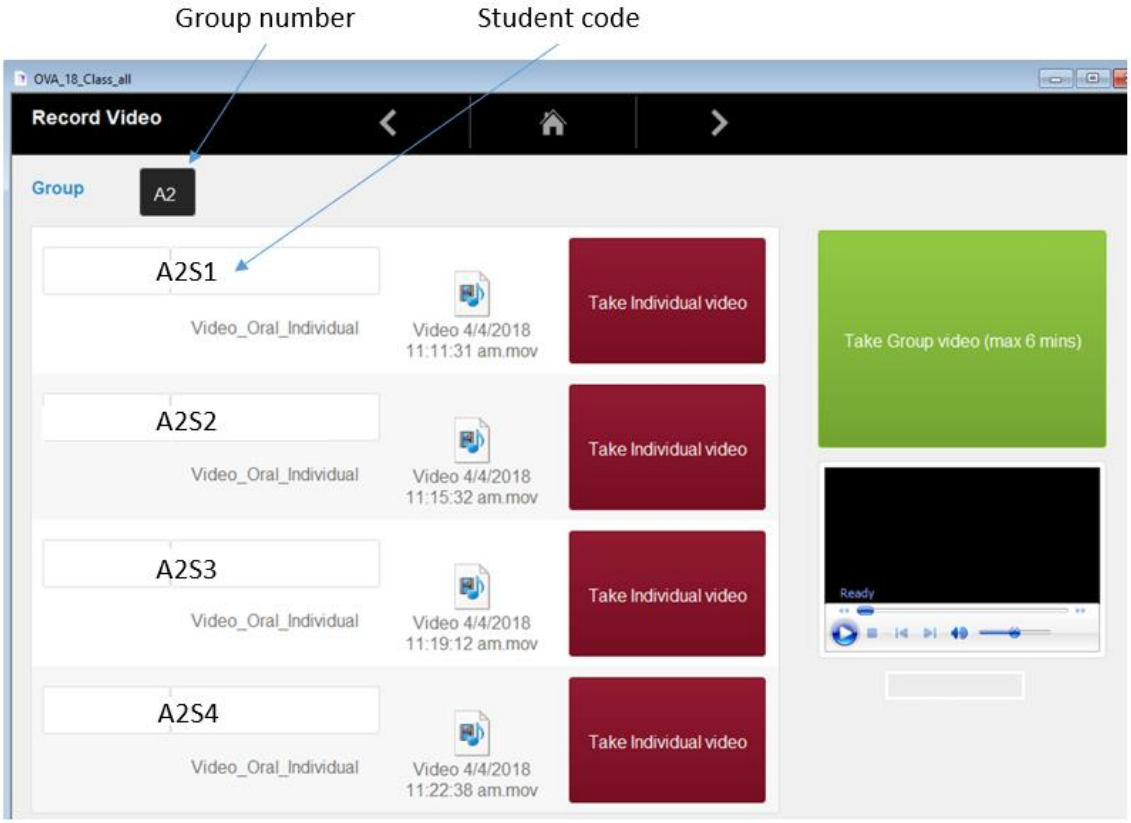


Figure 3.11 Video Recording Interface.

Students' names were coded to maintain confidentiality and contribute to objective marking. The name list was added to the App before videoing commenced and students were grouped randomly, regardless of gender or English competence. Teachers commenced recording by clicking the *Take Individual Video* button. Similarly, clicking the *Take Group Video* button started the video recordings of group performances. Group videos were prioritised to reduce the waiting time between assessments for students as much as possible.

Each recording function was allocated a set time – for individual videos the maximum time was three minutes, and for group videos, the maximum was six minutes. The time allowance was determined by the existing English speaking test at FPT University at the time of the research. Teachers could manually stop videoing if students finished their talks early, otherwise the recording stopped automatically when the set time limit was

reached. Student performances were automatically saved and stored in the App together with date, time and file format details.

Teachers were able to quickly and easily return to the home page by clicking on the Home button on the task bar at the top of the screen. Alongside the Home button, the Backward and Forward buttons allowed for toggling between screens, adding to the flexibility and practicality of the application.

Marking Function

Teachers had the option of marking offline on iPads or in the Cloud via a browser. Figure 3.14 shows the arrangement of videos in the marking interface. The OVA App catered for two speaking assessment tasks for each student: an individual and group assessment task, so there were two options for *Assessment Task Marking*: an individual task and a group task interface. The *Both Together* interface offered a time-saving option. The marking interface displayed student results for each assessment task and the total result for the two tasks; the latter calculated automatically when teachers imported the marks for each criterion in the marking key.

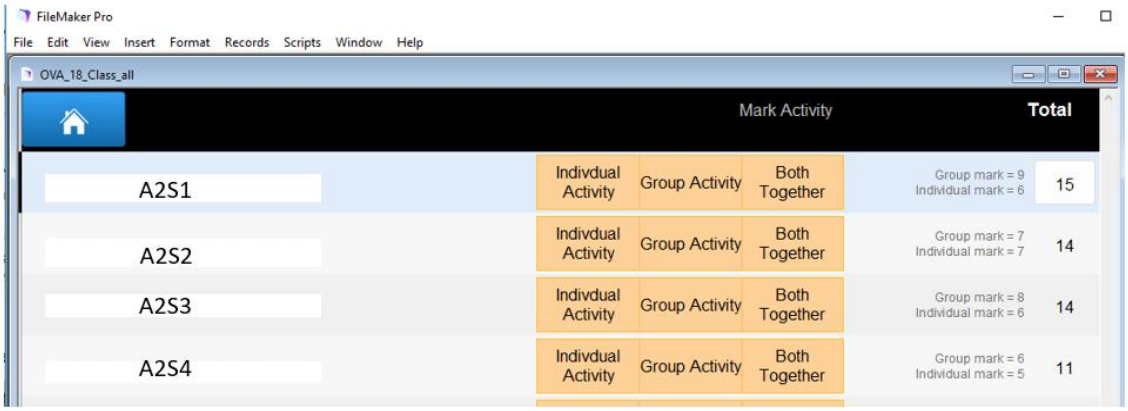


Figure 3.12 Marking Interface.

Selecting *Individual Activity* took teachers to the Individual Assessment Task Marking Interface (see Figure 3.13) containing the video of the student's individual task and the marking key for this task. The App allowed teachers to start, stop and replay the videos an unlimited number of times. Marking simply required clicking on each criterion of the marking key. For example, when marking fluency, teachers clicked on fluency criteria with three different levels from low to high. Fluency marks were added to the other criteria results marked in the same way and the total displayed at the bottom of the screen. In the bottom left corner, a small text box offered assessors an option to provide feedback.

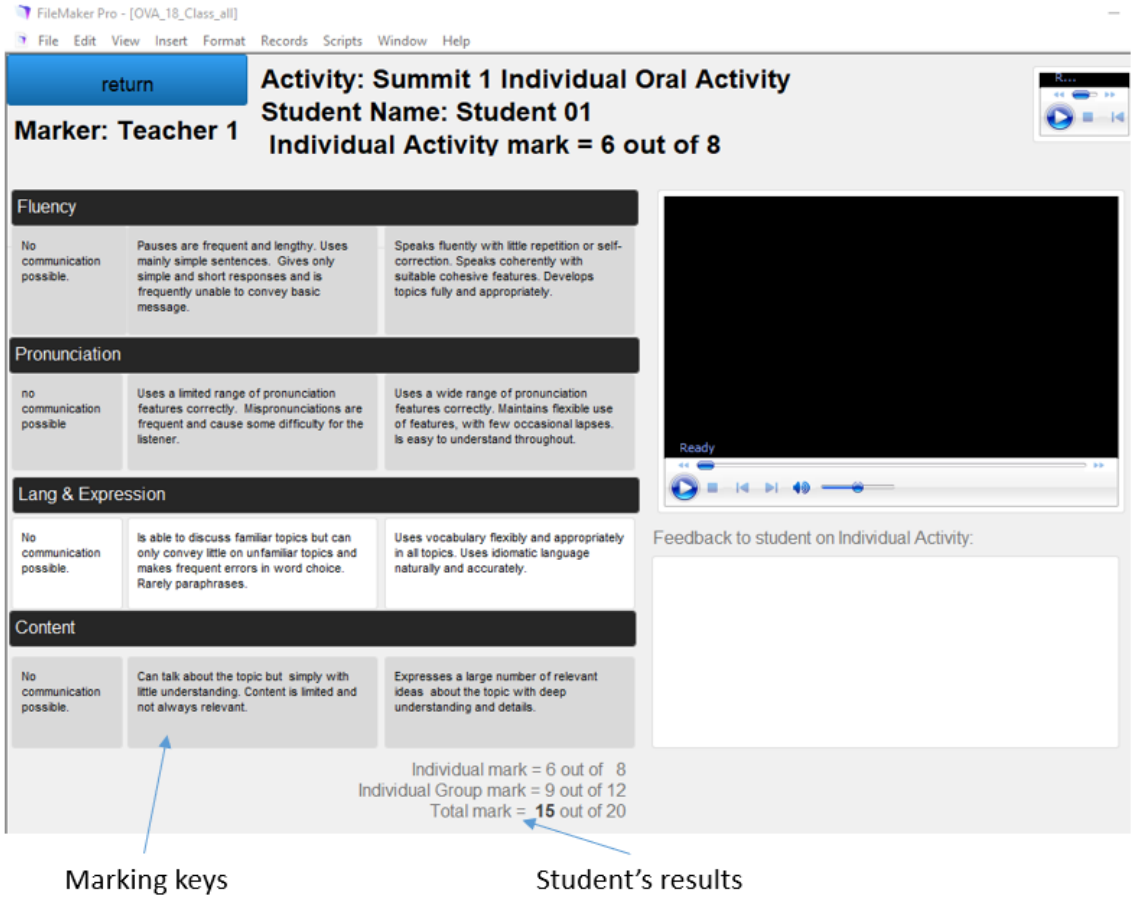


Figure 3.13 Individual Assessment Task Marking Interface.

Marking the group assessment task followed a similar pathway, with the exception of the marking key for the group task that contained four criteria, each weighted differently and some with more divisions than others (see Figure 3.14). In the same way as for individual tasks, teachers selected the relevant criteria. A photograph of the student was also provided to help teachers identify the individual within the group. Multiple marking and peer marking options were available by sharing videos and

multiple access to the Cloud. The App also facilitated moderation via email exchanges and discussion.

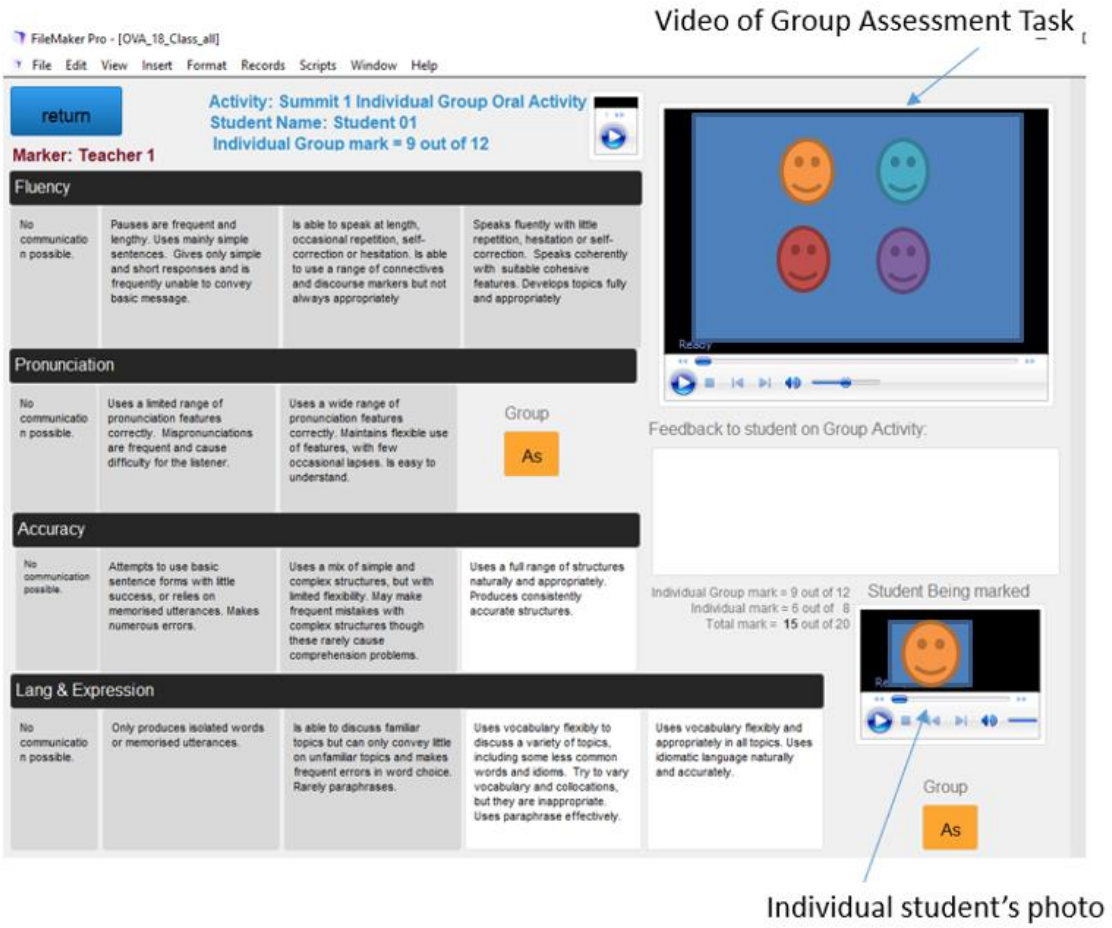


Figure 3.14 Group Assessment Task Marking Interface.

Managing Functions

Storage

The videos and results of student speaking performances were saved on iPads and in the Cloud for different purposes. Figure 3.15 shows how group results were arranged in the App, allowing for display of four individual results in one group task either by marker (see Figure 3.15) or by student, together with the results awarded by each marker (see Figure 3.16). This function assisted comparison among group members and teachers.

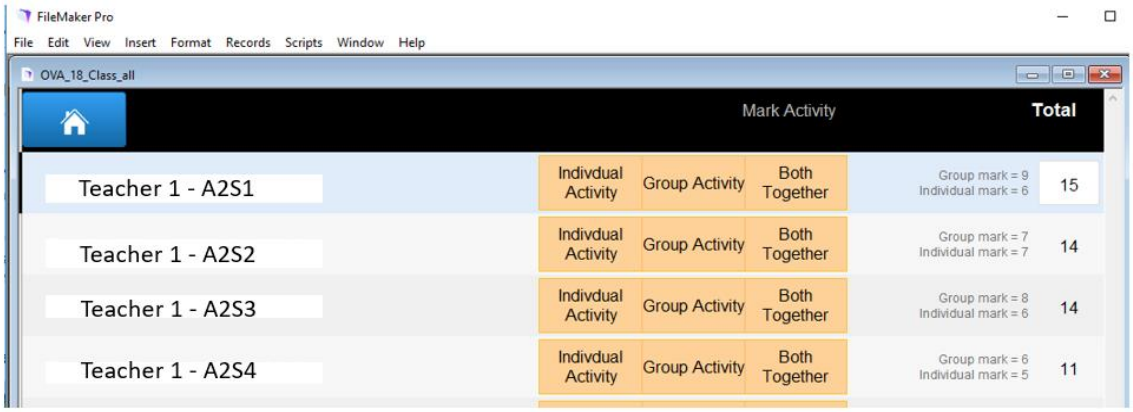


Figure 3.15 Group Marking Results.

Figure 3.16 shows how the results awarded by the different teachers were arranged in the App. This function facilitated moderation and multi-marking and allowed for measuring inter-rater reliability. It also fostered moderation, administration and review, as the differences in results from the different teachers were clearly evident.

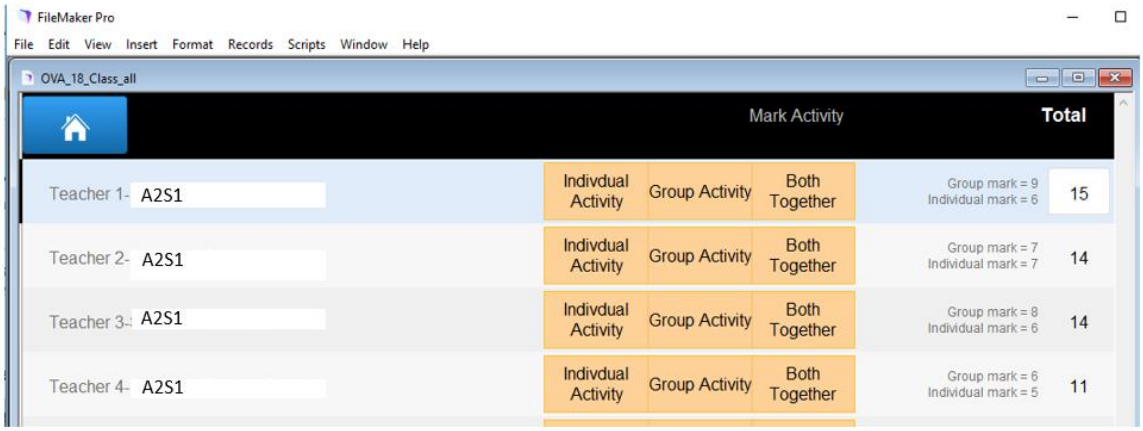


Figure 3.16 Multiple Marking Results.

Uploading and Sharing Activities

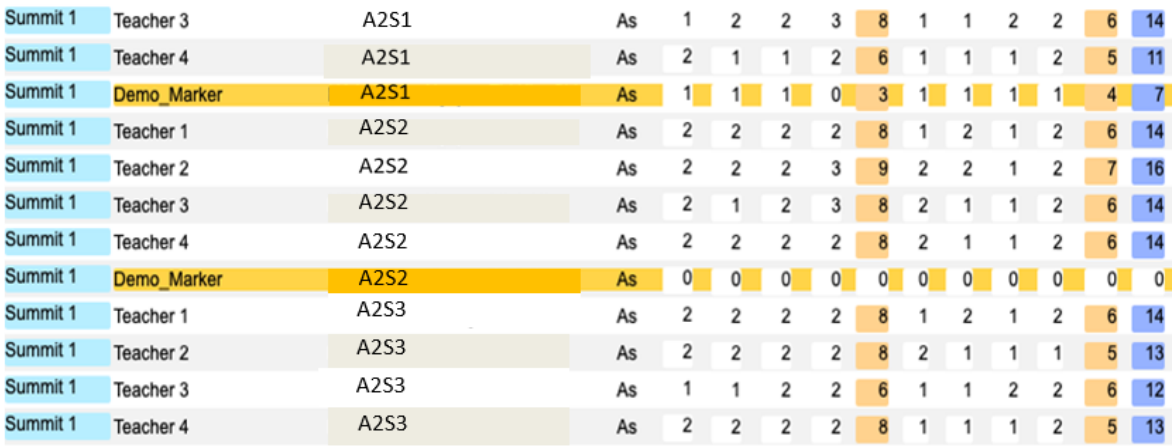
The OVA App allowed for videos to be seamlessly uploaded and stored in the application. Since the server was located in Australia and the students were in Vietnam, the decision was made to record the videos locally on an iPad. Teachers videoed the student performances on the App, and after recording an entire class of students, all the recordings were uploaded to the server. The administrator combined the data and uploaded the records to the Cloud.

Teachers and students were able to access the records via a Web browser. The administrator generated a user name and password for each teacher to log into the system and do their marking – all their marks and feedback were saved automatically.

Students could check their results and feedback using a computer or mobile device with internet connection or Wi-Fi access. Assigning unique usernames and passwords meant that teachers could manage the time and speed of their marking, edit the feedback and finalise the results before submitting.

Extracting and Reporting Results

The App had the capacity to export test results to Pdf files and Excel spreadsheets, where they could be sorted in alphabetical order by student names, by teacher or by group, depending on the requirements. Feedback on individual and group performances could be exported as Pdf files or Excel spreadsheets, and extracts of student results could be printed or emailed to teachers, students and administrative staff who distributed and archived the test results. Figure 3.17 shows an Excel spreadsheet of students' test results sorted by marker.



Summit 1	Teacher 3	A2S1	As	1	2	2	3	8	1	1	2	2	6	14
Summit 1	Teacher 4	A2S1	As	2	1	1	2	6	1	1	1	2	5	11
Summit 1	Demo_Marker	A2S1	As	1	1	1	0	3	1	1	1	1	4	7
Summit 1	Teacher 1	A2S2	As	2	2	2	2	8	1	2	1	2	6	14
Summit 1	Teacher 2	A2S2	As	2	2	2	3	9	2	2	1	2	7	16
Summit 1	Teacher 3	A2S2	As	2	1	2	3	8	2	1	1	2	6	14
Summit 1	Teacher 4	A2S2	As	2	2	2	2	8	2	1	1	2	6	14
Summit 1	Demo_Marker	A2S2	As	0	0	0	0	0	0	0	0	0	0	0
Summit 1	Teacher 1	A2S3	As	2	2	2	2	8	1	2	1	2	6	14
Summit 1	Teacher 2	A2S3	As	2	2	2	2	8	2	1	1	1	5	13
Summit 1	Teacher 3	A2S3	As	1	1	2	2	6	1	1	2	2	6	12
Summit 1	Teacher 4	A2S3	As	2	2	2	2	8	1	1	1	2	5	13

Figure 3.17 Test Results on an Excel Spreadsheet.

In conclusion, the OVA App functioned as a tool for collecting data and providing a digital environment for teachers to mark student speaking performances. It provided a platform for digital assessment to address the main research question in relation to manageability and functionality of the technology.

Ethical Considerations

The study participants comprised EFL students and teachers, aged between 18 and 55, at FPT University in Vietnam. There were no children involved in the research. The teachers were invited to participate by email and asked to email the information letter, consent form and invitation letters to their students (see Appendices C, D, Q, and R). All participants were recruited on a voluntary basis; they remained anonymous and

could withdraw from the research without penalty any time before the trial test in Phase 2. The video recordings were only used for marking and were presented in the thesis in a way that does not reveal the participants' identity. Participants were selected in order, as they volunteered, until the full quota was met, and could contact the researcher with any questions and concerns about the research.

Participants were provided with an information letter that clearly explained the research goals and the benefits of the research and highlighted any issues to consider before deciding to participate. They received consent letters via email, again with full disclosure of the nature, benefits and potential risks of the study. The information letter and consent letter were translated into Vietnamese so that they could fully understand the process.

The collected data were kept confidential, anonymous and used only for the purpose of this research. The audio and video recordings were only accessible to the teachers who did the marking, the researcher, and authorised supervisors from Edith Cowan University. The data is password protected and will be stored for five years after completion of the thesis, in compliance with The National Statement on Ethical Conduct in Human Research.

Summary

In summary, this chapter presented the methodology and mixed methods approach used to seek answers to the research questions investigating the feasibility of digital assessment for EFL speaking performance at tertiary level in Vietnam. The approach enabled triangulation of the different data sources, i.e., both quantitative and qualitative, to obtain an in-depth understanding of the phenomenon under study.

Phase 1 of the research explored participants' perceptions of using computer-assisted methods to assess EFL speaking skills at universities, their acceptance of this testing method, and willingness to attend a speaking trial using digital devices. Phase 1 informed Phase 2, which investigated the feasibility of a digital assessment method for student EFL speaking performances.

Various instruments were used to collect data for the study, including surveys, semi-structured interviews, observations and a trial test of EFL speaking skills. A customised tool, the OVA App, digitised the student performances, and assessments were undertaken and saved online. All the data were subjected to statistical analysis, NVivo theme coding, Cronbach's alpha reliability coefficient and Pearson correlation

coefficient analysis, in accordance with Kimbell et al.'s (2007) feasibility analysis framework. The mixed method design of the study served to validate the findings, provide an in-depth understanding of the research problem, and address the research questions, informed by an extensive review of the key literature.

The next chapter, Chapter 4, presents the findings of Phase 1 and proposes answers to research subquestion one: What are teacher and student perceptions of computer-assisted EFL speaking assessment?

CHAPTER 4
PHASE ONE FINDINGS

In Phase 1, data were collected via online surveys from two different groups of participants, university EFL teachers and students, to explore their perceptions of computer-assisted English speaking assessment. Their feedback was then analysed in relation to their willingness and acceptance to apply technologies for assessing EFL speaking skills. The findings of Phase 1 informed Phase 2 of the study.

A total of 278 (N(S1) = 278) students and 17 (N(T1) = 17) teachers responded to the surveys. They identified some important findings, presented in this chapter by group and according to emerging themes. Teacher perceptions are presented first, followed by student perceptions of computer-assisted EFL speaking assessment. Tables and graphs demonstrate statistical data and clarify the findings.

Teacher Perceptions

Teacher Demographic Information

There were 17 teacher participants, 14 females and three males, most (10/17) in the 35 to 44 age range. The majority (15/17) had over five years' experience teaching EFL. The survey data showed that all teachers (17/17) used laptops to support their teaching, many used smartphones (10/17), and some used desktop computers (5/17), and tablets (3/17) for teaching English.

Computer-Assisted EFL Tests

The data showed that computer-assisted English tests were frequently used by the teachers. They included existing and customised, teacher-designed online tests, automatically scored online tests, and tests taken by students on computers and then downloaded and marked by teachers.

Analysis revealed a dominance of computer-assisted English tests in the classrooms under study. Sixteen (16/17) teachers used online or computer-assisted tests, fifteen (15/17) claimed they used speaking tests, and nine (9/17) used paper-and-pencil tests. Computer-assisted tests were used more frequently than paper-and-pencil tests and oral tests. The English testing techniques used are shown in Figure 4.1.

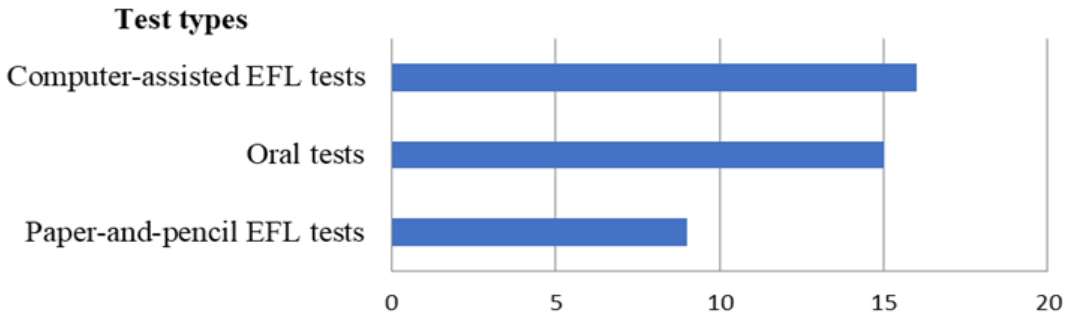


Figure 4.1 Frequency of Test Types used in EFL Classrooms.

Eight out of seventeen (8/17) English teachers had attended training courses to design, customise and deliver computer-assisted English tests. Most of the courses provided them with knowledge and skills to use the university's CMS (Content Management System), an internal website for university teachers and students to deliver tests and access learning materials. They also received training in Moodle, Testmoz, and Quizizz, websites and applications for generating online-delivered tests. In addition, teachers attended periodical training courses at the university to learn how to build online test databases using the internal website (CMS). The indications were that teachers were knowledgeable about certain specific test-generating websites and applications.

Most teachers (9/17) were familiar with and used online tests available from websites such as www.ego4u.com,, www.learnrealenglish.com, www.Englishexercises.org, www.takeielts.bristishcouncil.org, and www.Englishaula.com. More than 75% of the teachers (13/17) used websites and online tools to design their own tests, having obtained most of the tools from university training courses, such as CMS, Moodle, Testmoz, and Quizizz. Some teachers also used Kahoot, Quizlet, and Quia to design and deliver tests. The data indicated that a high proportion of teachers (13/17) were familiar with English testing websites and had experience adapting and designing their own online tests to suit their specific purposes. They were also capable of integrating technologies to enhance their test practice. Teachers expressed a preference for computer-assisted tests and were evidently competent in the use of IT for test design and delivery.

Most of the teachers (9/17) surveyed had minimised their use of paper and pencils for tests. As shown in Figure 4.1, paper-and-pencil tests were the least used compared to oral and computer-assisted tests.

EFL Speaking Tests

Fifteen (15) teachers claimed they used live speaking tests to assess students' English proficiency. They ranked second in terms of popularity compared to the other two forms of testing. The data suggested that integrated computer assistance would benefit students and save teachers time.

Computer-Assisted EFL Speaking Tests

The data showed that all 17 teachers (17/17) surveyed used computer-assisted tests to evaluate students' reading skills; sixteen (16/17) used them frequently for assessing students' listening skills. Some teachers designed online tests for writing skills (6/17), grammar and vocabulary (4/17). Only two teachers (2/17) reported using computer-assisted tests to evaluate speaking skills. Figure 4.2 shows the frequency of use for computer-assisted tests across all language skills.

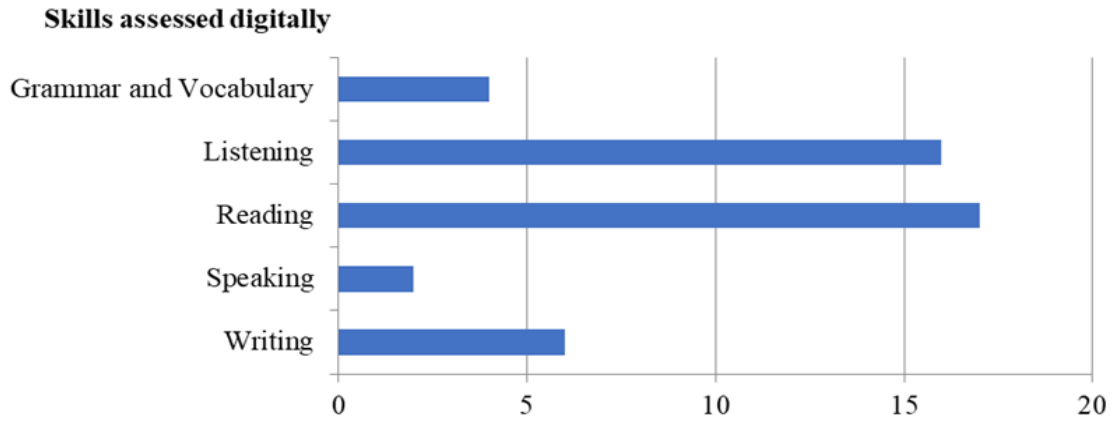


Figure 4.2 The Use of Computer-Assisted Tests for Each English Skill.

The numbers show that computer-assisted tests were used infrequently for speaking skills. This could be attributed to the difficulties of integrating technologies into speaking tests or a lack of training among teachers to design such tests on computer. It may also be possible that internet websites and tools did not support online testing of English speaking skills or teachers had difficulties accessing available online computer-assisted speaking tests.

Teacher Preferences

Most teachers (15/17) indicated a preference for computer-assisted English tests to assess students' proficiency. This was consistent with the number of teachers who chose computer-assisted tests for assessing students' English competence (see Figure 4.1).

Teachers' perceptions of the current paper-and-pencil testing method revealed that most (14/17) found it time-consuming and expensive. The majority (11/17) believed that it was reliable, and eight (8/17) teachers considered it fair. Few teachers (2/17) agreed that this testing method was authentic, objective and easy to manage, and all of them identified the lack of immediate feedback and interaction in the paper-and-pencil method as drawbacks. Figure 4.3 shows the differences in teachers' perceptions of paper-and-pencil and computer-assisted tests.

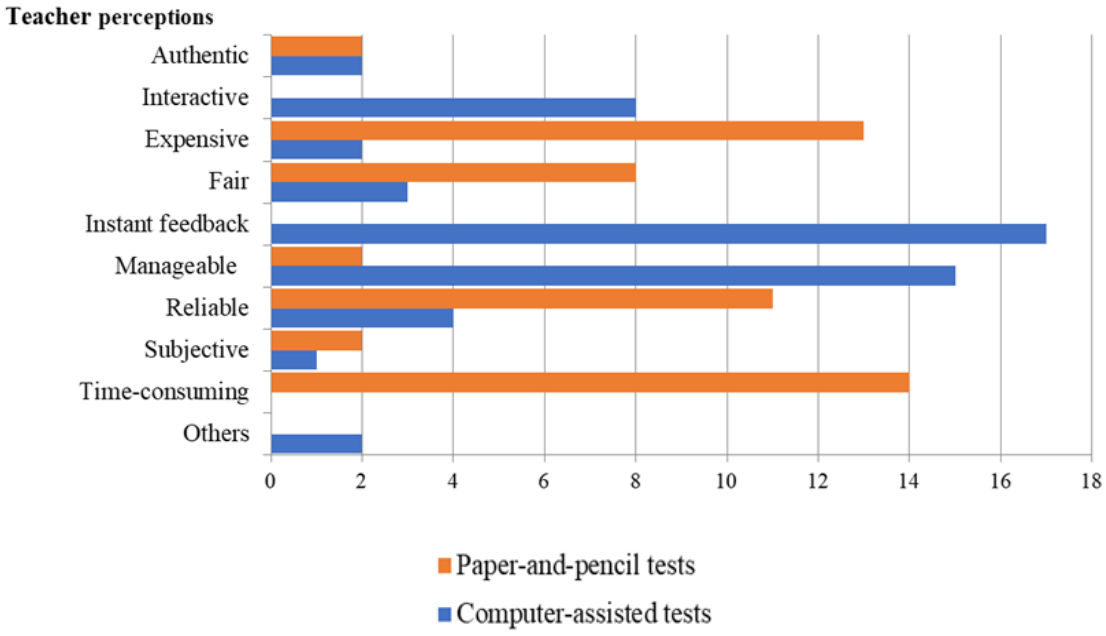


Figure 4.3 Teacher Perceptions of EFL Assessment Methods.

Teachers (17/17) all agreed that computer-assisted EFL tests provided students with more immediate feedback. Compared to paper-and-pencil tests, many teachers (15/17) found computer-assisted tests manageable, and eight (8/17) believed it offered more interaction. Four (4/17) teachers considered the digital testing method reliable, three thought it was fair, and two found it authentic. Few thought it was expensive (2/17) and subjective (1/17), and none of the teachers viewed it as a time-consuming method. This data indicated that most teachers thought subjectivity in scoring and the financial costs of using computer-assisted tests were an issue. Most believed that the digital testing method could provide instant feedback to both teachers and students and facilitated test administration. In addition to immediate feedback, teachers were positive about the advantages of computer-assisted English tests, including their manageability, objectivity, time and financial efficiencies. Two teachers commented on the interfaces of computer-assisted tests as being easy to edit and update, saving time and costs.

Overall, teachers were somewhat cynical about the reliability and authenticity of digital tests. Only four (4/17) considered them reliable and two (2/17) found them authentic. Their scepticism may be due to their lack of experience in choosing reliable online exam resources and the way in which they delivered tests to their students.

In summary, the surveyed teachers had a preference for computer-assisted English tests over the current paper-and-pencil tests, and perceived computer-assisted tests offered more advantages in terms of feedback, manageability, time and costs. This perception appeared to underpin the popularity of computer-assisted tests in English classes and had led to a reduction of paper tests in practice.

Teacher Experience

Teacher participants were provided with a clear definition of computer-assisted EFL speaking assessment before they completed the survey. The concept covered all speaking tests supported by computers and other digital technologies with additional functions, ranging from video and audio recordings to automated scoring and feedback generation. Thirteen (13/17) teachers had never before delivered any computer-assisted speaking tests with video and audio recording. Twelve (12/17) teachers used face-to-face interviews to assess their students' speaking skills. A few (3/17) indicated they used computers for speaking tests and retained video and audio recordings of the performances. Two teachers (2/17) described their students speaking as monologues, while they listened from beginning to end without asking any questions or providing any feedback.

Face-to-Face Interviews

The data showed that face-to-face or direct interviews were frequently used to assess students' speaking competence. Twelve (12/17) teachers claimed they used this method over any others. Many agreed that face-to-face interviews offered interaction (13/17) and authenticity (11/17). Eleven (11/17) considered face-to-face interviews to be reliable, and nine (9/17) concurred that it facilitated instant feedback.

More than half the teachers (11/17) found organising interviews time consuming and nearly half (8/17) had concerns about subjectivity associated with this method. The majority (15/17) believed that interviews were difficult to manage. Only three teachers (3/17) made recordings of student oral performances for later review, while they assessed students' speaking skills in face-to-face interviews. Figure 4.4 shows the

differences in teacher perceptions of face-to-face interviews and computer-assisted speaking assessments.

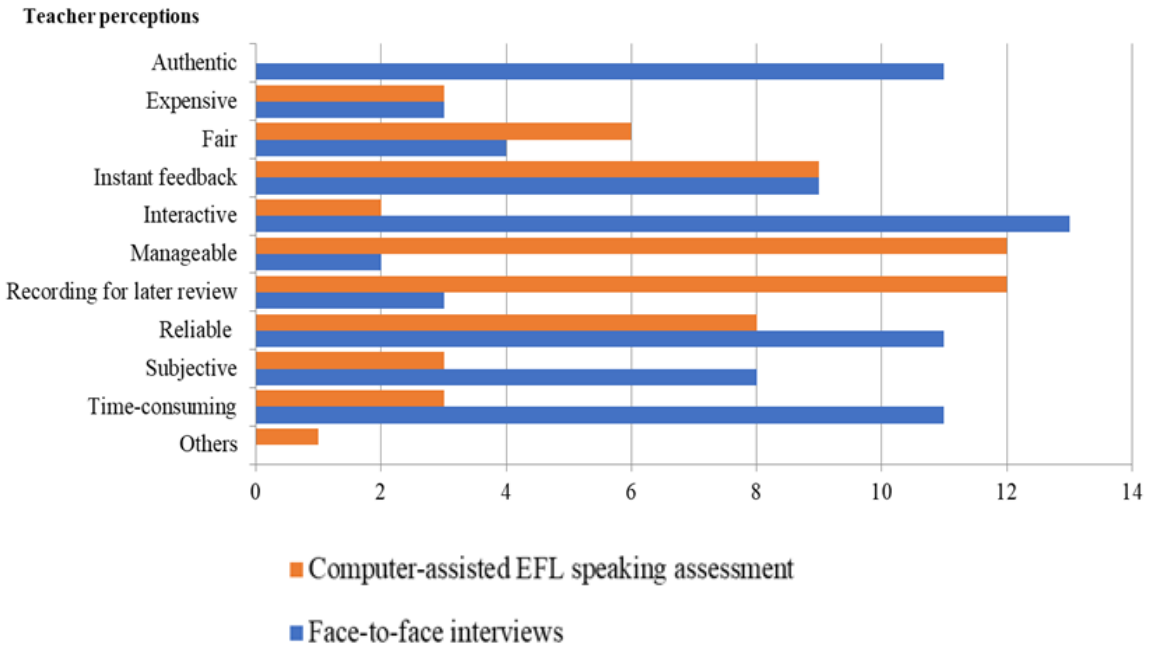


Figure 4.4 Teacher Perceptions of EFL Speaking Assessment Methods.

Teacher Beliefs about Digital Assessment

The data showed the majority of teachers perceived computer-assisted speaking assessment offered easier test administration (12/17) and recognised the benefits of recording student performances for later review (12/17) compared to face-to-face interviews. They also agreed that computer-assisted speaking assessment significantly reduced the time and subjectivity in scoring and argued that digital assessment could provide as much immediate feedback and interaction as face-to-face interviews. However, they were sceptical about the reliability of digital testing and doubtful that it could offer as much authenticity as interviews. This could be attributed to their lack of hands-on experience with computer-assisted assessment and signalled the need for a digital test trial.

Based on the survey data, the biggest differences in teacher perceptions of face-to-face interviews and computer-assisted speaking assessment were in areas of interaction, time, authenticity and recordings of tests for later review. On the one hand, they believed that face-to-face interviews involved significant interaction between teachers and students and were more authentic in imitating real-life contexts. On the other hand, the majority of teachers (11/17) found interviews time-consuming, and in the absence of recordings, lacked test evidence and therefore capacity for later review.

Computer-assisted speaking assessment was considered to be time efficient and easy to manage. The recordings of students' speaking performances provided test evidence and opportunities for later review. It was seen as a less subjective and fairer method of scoring student performances. Teachers commented that it was a modern, progressive and professional way of conducting speaking tests.

The advantages of computer-assisted EFL speaking assessment were perceived to outnumber the benefits of face-to-face interviews. Although interviews were considered more reliable, they were also more subjective, time-consuming and difficult to manage. Nearly half the teachers (7/17) expressed a preference for computer-assisted assessment over face-to-face interviews because the digital approach offered time efficiency and manageability. A third (6/17) were cynical about the reliability of the digital method and lacked the confidence to use it as a replacement for conventional interviews.

Perceived Usefulness and Ease of Use

Nine constructs were used to describe Perceived Usefulness (U) from the perspectives of teachers, with eight out of nine (8/9) identified. Teachers perceived computer-assisted assessment useful, both educationally and economically. They believed it improved the reliability of speaking tests, provided immediate feedback, reduced subjectivity, and enhanced fairness. In terms of cost, computer-assisted assessment lowered the demand on time and facilitated test management. Table 4.1 shows a list of Perceived Usefulness constructs and the survey results.

Table 4.1

Teacher Perceptions of Perceived Usefulness Constructs

Items	**Perceived Usefulness**	**Results**
U1	Enhancing fairness	35% (6/17)
U2	Facilitating exam administration	71% (12/17)
U3	Improving the reliability of English speaking tests	47% (8/17)
U4	Offering authenticity	0% (0/17)
U5	Offering better interaction compared to face-to-face interviews	12% (2/17)
U6	Providing immediate feedback	53% (9/17)
U7	Reducing subjectivity in rating students	82% (14/17)
U8	Saving financial costs	82% (14/17)
U9	Saving time	82% (14/17)

Adapted from F. Davis (1989)

The survey results showed that items U2, U7, U8, and U9 received the most positive responses. More than 50% of the teachers surveyed agreed most frequently on items U7, U8, and U9, indicating that computer-assisted assessment was strongly believed to

be efficient in terms of time, cost and objectivity in scoring. Item U4 (management) was also agreed by 12 out of 17 teachers.

Four (4) constructs were used to describe Perceived Ease of Use (E), with three out of four (3/4) identified: (a) providing recordings of student speaking performances for later review, (b) an easy-to-use interface, and (c) reducing stress and nervousness. Table 4.2 presents the survey results for Perceived Ease of Use constructs.

Table 4.2

Teacher Perceptions of Perceived Ease of Use Constructs

Items	Perceived Ease of Use	Results
E1	Giving convenience in terms of test time and test locations	6% (1/17)
E2	Offering easy-to-use interfaces	6% (1/17)
E3	Providing recordings for later review	71% (12/17)
E4	Reducing stress and nervousness	0% (0/17)

Adapted from F. Davis et al. (1989)

Item E3 (recordings for later review) received the most agreement amongst teachers (12/17). Most believed that computer-assisted assessment could facilitate review of student performances through the use of audio and video recordings. One respondent's reference to computer-assisted assessment being professional and modern was coded E2 (offering easy-to-use interfaces). A further comment was coded E1 (convenience in terms of test time and test locations) in reference to digital assessment saving teachers time. No responses were coded to E4 (reducing stress and nervousness), possibly an indication that this issue wasn't as relevant.

In summary, both Perceived Usefulness and Perceived Ease of Use were identified and indicated that teachers had positive perceptions of computer-assisted assessment in terms of these constructs.

Teacher Acceptance of a Speaking Test Trial

Although the teachers had different views about computer-assisted EFL speaking assessment, the majority (11/17) expressed strong acceptance of a computer-assisted speaking trial. A third of them (4/17) were cynical, and two declined to participate, claiming that it was "not authentic interaction" (Q22 – Teacher Survey responses). Figure 4.5 shows the teachers' acceptance of a computer-assisted EFL speaking trial.

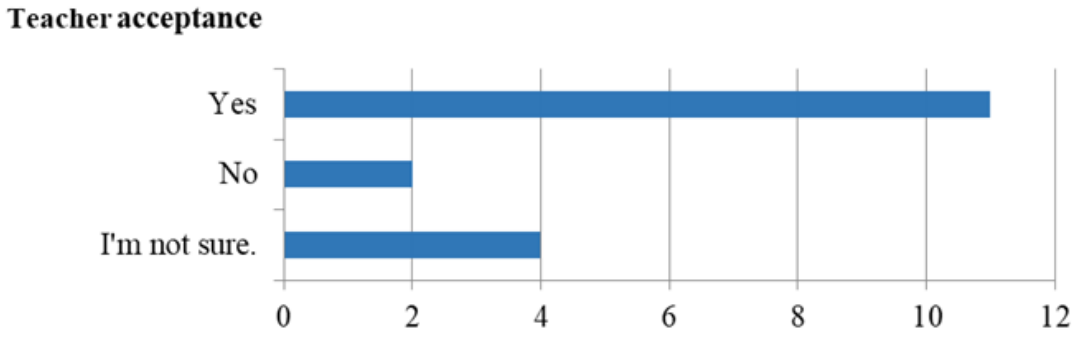


Figure 4.5 Teachers' Acceptance of a Trial.

Based on the technology acceptance model (F. Davis et al., 1989), most teachers had a positive attitude towards the digital testing approach. The introduction of a computer-assisted speaking trial was deemed appropriate to strengthen the research findings in Phase 2 and further examine the feasibility of computer-assisted EFL speaking assessment in the Vietnamese context.

Student Perceptions

Student English and ICT Literacy

A total of 278 university EFL students (N(S1) = 278) responded to the survey: 81% were male and 19% female. Their English competency ranged from beginner to advanced level. Of the cohort, 29% had intermediate English, and only 4% possessed advanced English, with most students at pre-intermediate level and lower.

Ninety-six percent of the students had laptops and 76% possessed smartphones as study resources. Eighty-two percent used digital equipment every day to support their English learning. Facebook was the most popular website, accessed by 70% of students for study. Nearly 50% of students used English learning websites and 39% used Google Docs to learn English. A large number of other websites were mentioned as regular sources for language learning; among them Quizlet, Doulingo and Youtube were most popular and Quizlet enjoyed the highest user rate. Students also indicated that they used a large number of online dictionaries, such as online Oxford dictionaries (Oxforddictionaries.com), online Cambridge dictionaries (Dictionary.cambridge.org), and Vdict (7.vndic.net and Vdict.com). Many used online testing websites, such as Englishteststore.net, Englishaula.com, and Quizizz.com. It was evident from the survey results that students were familiar and confident with online EFL learning and testing programs. In addition, students accessed applications that helped them learn to speak English like native speakers. The most popular of these was English Language Speech

Assistant (ELSA), an application for mobile phones that provides language learners with instant feedback on pronunciation, assessment tests and lessons designed by pronunciation experts. The application can be downloaded from www.elsanow.io.

In summary, students had full access to modern technology and high levels of IT literacy. Data obtained from the initial survey indicated that students were already using online tools and websites to improve their English speaking skills, so computer-assisted EFL assessment was not unfamiliar to them.

Computer-Assisted EFL Tests

According to the data, all students took English tests at the end of each semester; the majority of these computer-assisted. Approximately 45% of students said they took computer-assisted English tests. A smaller number of speaking tests used the paper-and-pencil method. This is consistent with the survey findings on teachers' use of computer-assisted English tests in their practice. Figure 4.6 shows the distribution of trends for the different types of tests in English classes.

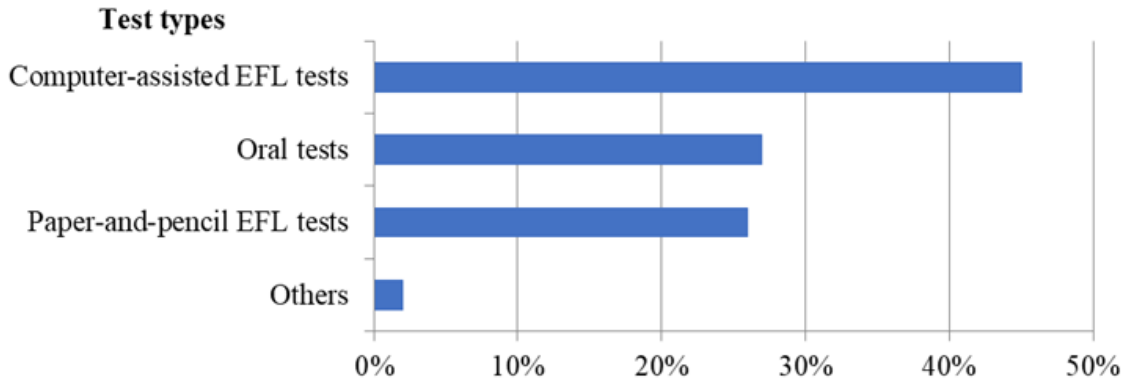


Figure 4.6 Types of Tests Taken by Students in English Class.

Student Preferences

More than 70% of students said they preferred computer-assisted tests over paper-and-pencil tests and oral tests. Over 15% claimed that they liked oral tests, and 14% said they liked the current paper-and-pencil tests. Figure 4.7 shows students' preferences for the different types of English tests.

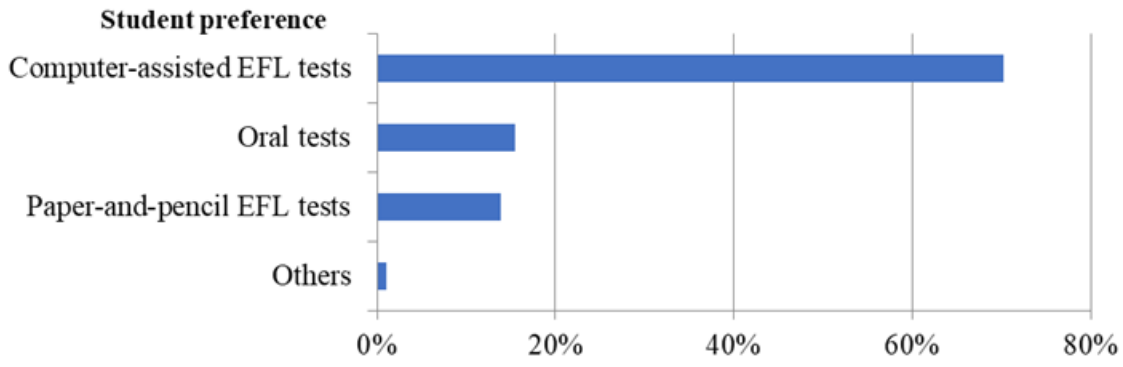


Figure 4.7 Student Preferences for Different Types of Tests.

The students had different reasons for preferring computer-assisted tests; the most common one was the convenience they offered. They could be completed at any time and in any location. "Convenient" was the most frequent response. A large number of students agreed that the ability of computer-assisted EFL tests to provide instant results and feedback was also a benefit. "Fast"; "immediate results, instant reports of test results"; "the results are correct and announced to students fast"; and "save time" were all common responses. Students found interacting with the test interface easy and user-friendly, and admitted not having to worry about their bad handwriting.

Students credited digital testing with offering access to a broad range of test questions and being a paper-saving strategy. Stress reduction was another motivation for their interest in this type of test. Some mentioned "reducing our stress" and "fun" to describe their thoughts in relation to computer-assisted English tests. They believed that interacting with a computer was far more relaxing than sitting in front of an examiner in a face-to-face interview.

Although the majority of students regarded computer-assisted EFL tests as "professional" and "modern", a few were concerned about security. They were worried about how this testing method would prevent cheating and mitigate against random choosing of answers.

Although computer-assisted tests were preferred by most students, the other two testing methods were also viewed as effective and beneficial. Fourteen percent of students preferred paper-and-pencil tests because they were unfamiliar with computers and lacked typing skills. Students said: "Because I love using pencils" and "I'm not good at technology". They were more confident with paper tests because they could write down draft answers and review them before submitting. They said: "Having tests on the paper is easy to read question and write the answer". Some students claimed the paper tests

helped them better memorise the content. Others refused to use computer-assisted tests because they were concerned about unexpected technical problems, such as internet disconnection and test submission failure, that could affect their test results. One student said: “Computers are sometimes disconnected from the internet, which directly affects students’ test results and other things. Paper tests do not have such issues”.

Approximately 16% of the student cohort indicated a preference for oral EFL tests, i.e., face-to-face interviews with one or two examiners and individuals or groups of three or four students. They believed that face-to-face interviews enhanced teacher-student interaction and the more interaction students were exposed to, the better their communication skills would become. Most students also believed that interviews provided them with opportunities to improve their pronunciation and listening skills from interviewers with different accents. Another reason offered was that interviews involved more authentic, real-life situations. Some students claimed that oral tests could easily and precisely assess their speaking competence. Others believed that oral tests enhanced their “soft skills”, such as negotiation, eye contact and facial expressions, all of which contributed to conversation.

Student Experience

The survey data indicated that computer-assisted tests were mostly used to assess reading, listening and writing skills, with speaking skills infrequently tested this way. Sixty-seven percent of students had their EFL reading, listening and writing skills tested by computer. Fewer than 20% had ever taken a computer-assisted speaking test (see Figure 4.8).

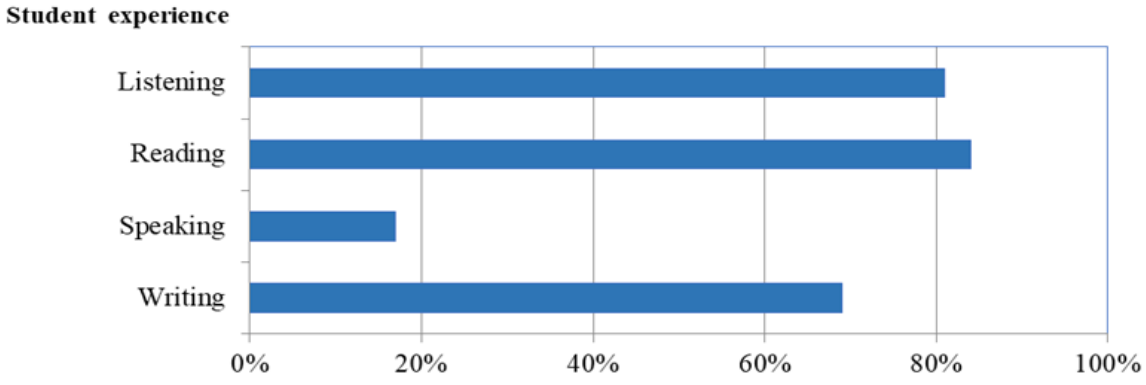


Figure 4.8 Student Experience with Computer-Assisted EFL Tests.

The majority of students (69%) surveyed expressed a preference for computer-assisted listening tests. Both computer-assisted listening and writing tests were preferred by over

60% of students, while a substantial number (26%) preferred speaking tests. This was higher than the number of students who had undertaken computer-assisted speaking tests (see Figure 4.9).

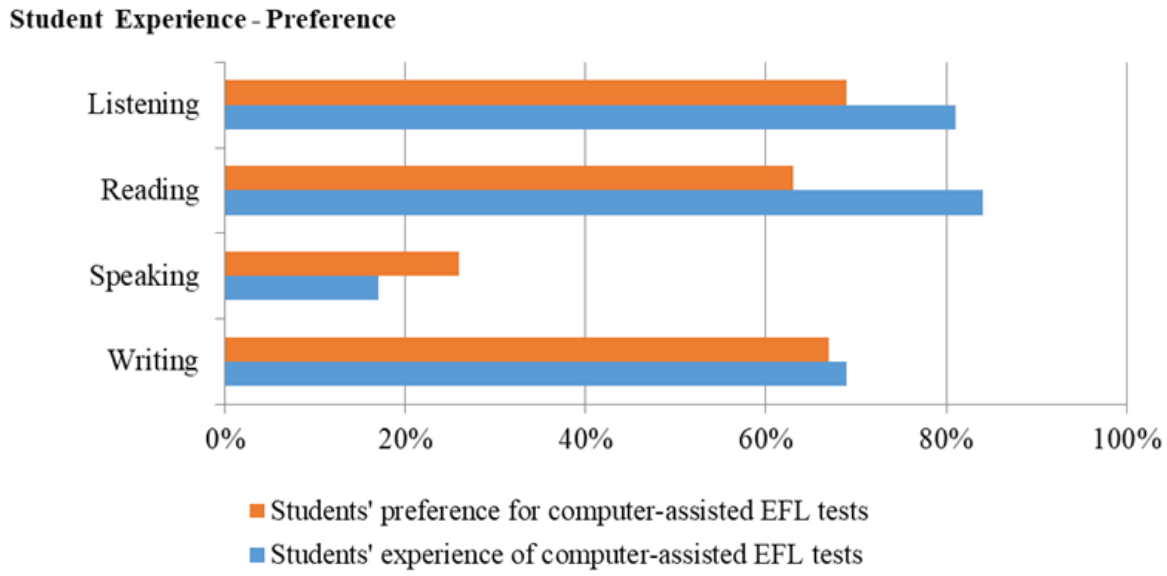


Figure 4.9 Student Experience and Preference for Computer-Assisted EFL Tests.

The discrepancy between actual use of computer-assisted English speaking tests and student preferences for this kind of assessment flagged demand and suggested that the practice of computer-assisted EFL speaking tests should be expanded.

Absence of ICT in Assessing EFL Speaking

The survey data indicated that face-to-face interview tests consisted of one or more speaking tasks, including face-to-face teacher and student interviews, group discussions with examiners observing and judging, speaking to a computer with audio and video recording, and face-to-face interviews with audio recording. Table 4.3 shows the frequency of each assessment task.

The most common testing activity was face-to-face teacher-student interviews (66%), followed by group discussions with examiners observing and judging (62%). The combined total of individual interviews and group discussions accounted for 59% of the overall mark, while other activities, such as speaking into a computer with audio and video recording and face-to-face interviews with audio recording were rarely used. Audio and video recordings were not used in English speaking tests at FPT University.

Table 4.3

English Speaking Assessment Tasks and Frequency of Use

Speaking tasks	**Frequency of use**
Both individual interviews and group discussion	59%
Face-to-face interviews with audio recording	5%
Face-to-face teacher student interviews	66%
Group discussion with examiners' observation and judgement	62%
Speaking to a computer with audio and video recording	12%
Others	3%

Student Perceptions of Speaking Assessments

The majority of student participants (66%) agreed that face-to-face interviews facilitated interaction between test takers and examiners. Forty-two percent stated that interviews were more authentic because the situations were similar to real-life contexts and conversations closely mimicked real-life communication. Some students complained that interview topics were sometimes unrealistic and unfamiliar to them. One student commented: "Unrealistic: Such as some speaking tests just ask about a subject that you don't know and it may make your test isn't good because you have to think a lot about that subject". For example, intermediate students (Top Notch 3) could be asked to talk about topics like "formal dinner etiquette", "comics: trash or treasure?", and "natural disasters" (Allen & Joan, 2011).

Thirty-seven percent of students said they received immediate feedback in face-to-face interviews, suggesting that examiners did not always provide feedback in the speaking tests and that some students got feedback while others did not.

Most of the students surveyed believed the existing testing method was reliable and fair – only 1% considered it unreliable and 3%, unfair. Overall, this method was viewed as being effective, since only a handful of students responded that it was subjective (10%) and time consuming (2%). Figure 4.10 shows the student perceptions of face-to-face interviews in English speaking tests.

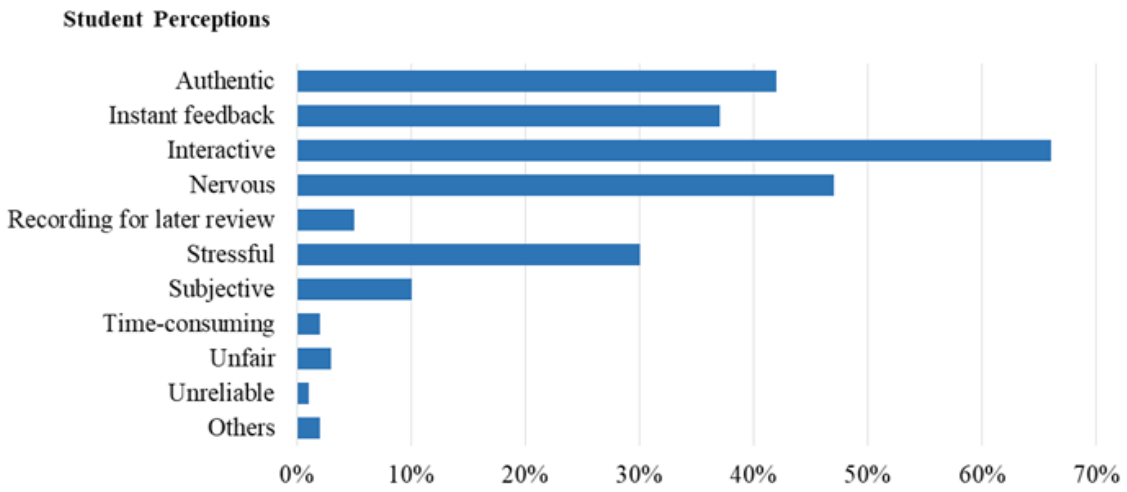


Figure 4.10 Student Perceptions of Speaking Assessments.

The students reported high levels of stress and nervousness in the survey. Nearly 47% stated they felt unduly nervous about face-to-face interviews with examiners and 30% said they felt stressed. A small number of students (12%) found face-to-face testing subjective, citing unfairness as an issue. Only 5% of the students were recorded for later review of their performances. The data suggested that student performances were primarily evaluated at the time of testing, without any recordings to provide test evidence for later review.

In summary, from the student perspectives, key issues were nervousness and stress about direct interviews in speaking tests. For them, the most positive aspect of face-to-face interviews was high levels of interaction and authenticity.

Computer-Assisted EFL Speaking Assessment Trial

Nearly three quarters (71%) of the students disclosed in the survey that they had never before taken an English speaking test in a digital format. However, when asked whether they thought computer-assisted speaking tests with audio and video recordings were a good idea, 55% agreed. Some students believed this approach would save time, reduce their stress levels, and eliminate subjectivity in scoring. They also recognised the benefits of being able to record their performances as evidence of their tests and for later review. Figure 4.11 shows student perceptions of computer-assisted EFL speaking assessment.

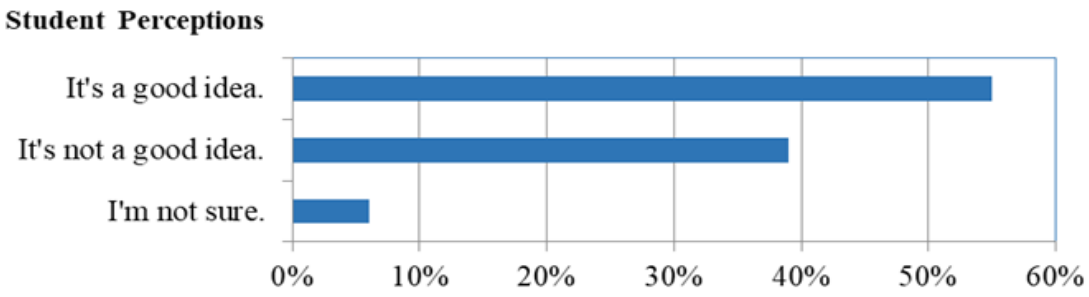


Figure 4.11 Student Perceptions of Digital Speaking Assessments.

Some students were sceptical about the digital method. In their opinion, it offered both advantages and disadvantages. Disadvantages were its dependence on technology and lack of authenticity because students talked to a computer, not a human examiner. They were concerned about their recorded voices not sounding natural, and that the technology could affect their performance. This accounted for 67% of students who preferred face-to-face interviews over the digital method for speaking tests (see Figure 4.12).

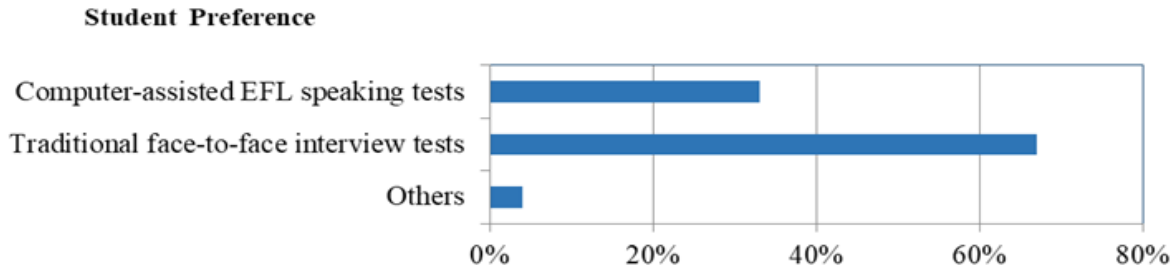


Figure 4.12 Student Preferences for EFL Speaking Test Methods.

Student Acceptance of the Speaking Test Trial

Figure 4.13 shows student acceptance of a trial computer-assisted EFL speaking test. More than 40% agreed to participate and forty-seven percent declined. Twelve percent weren't sure and asked to be contacted again later.

Figure 4.11 shows most students had a positive attitude towards the digital testing method. The number of those who thought computer-assisted EFL speaking assessment was a good idea was larger than the number who agreed to take part in the trial test, suggesting that students were sceptical about the new method in practice. According to the survey results, most students had no experience of taking a computer-assisted EFL speaking test; providing an opportunity to try the new testing method and see whether it changed their perspectives was a valuable prospect.

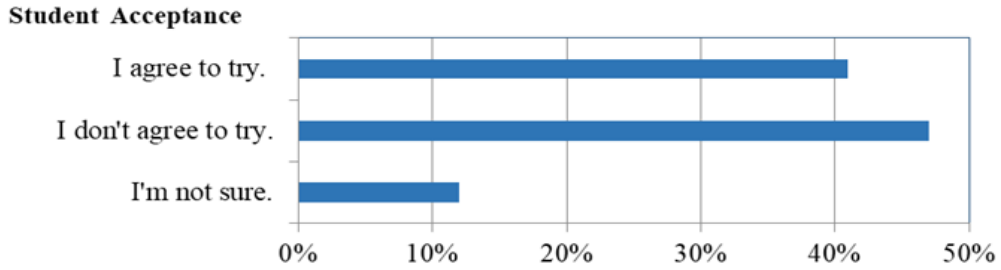


Figure 4.13 Student Acceptance of a Speaking Test Trial.

A comparison between acceptance of the trial test among teachers (see Figure 4.5) and students (see Figure 4.13) showed stronger interest from teachers. Both groups had some degree of doubt about digital assessment, reinforcing the usefulness of a trial test to determine its feasibility in real testing situations, further explore the views of users, and determine the implications for English speaking assessment.

Summary

The findings of this study supported strong acceptance of computer-assisted EFL speaking assessment by both teachers and students and underscored the potential value of introducing this method in a real testing situation. A trial would provide teachers and students with hands-on experience of the digital testing method, enhance their knowledge of computer-assisted language assessment, and promote the testing of English speaking.

Although computer-assisted speaking assessments had not previously been used by teachers and students in Vietnam, it had been proven feasible in other studies (Kimbell, 2012b; Kimbell et al., 2007; Newhouse & Cooper, 2013; Newhouse et al., 2011; Stables & Kimbell, 2007; Williams & Newhouse, 2013). The aforementioned explorations showed that computer-assisted speaking assessments reduced time and subjectivity and enhanced the reliability of speaking tests. The findings of the current study suggested that an initial trial of computer-assisted EFL speaking tests in some language classes at FPT university would be valuable under the following conditions:

- Language classes had laptops and internet access,
- Students and teachers had some knowledge and experience with computer-assisted language assessment,
- Teachers and students had high levels of Information Technology literacy,
- Teachers and students were willing, eager and accepting of the digital testing approach,

- There was an available IT system for computer-assisted language assessment,
- There was a need for a new testing method to improve testing quality and save resources.

Phase 1 was a preliminary study for the second phase of the research. It served to identify favourable conditions for introducing the digital testing approach, indicated potential risks, and provided demographic information about the participants in Phase 2. The findings of Phase 1 restated the need for Phase 2 to examine the feasibility of computer-assisted EFL speaking assessment in a real testing situation and further explore the views of users in a Vietnamese context.

CHAPTER 5
PHASE TWO FINDINGS

The previous chapter discussed student and teacher perceptions of computer-assisted EFL speaking assessment and their willingness to participate in a digital speaking test. It also examined the feasibility of digital speaking assessments using the OVA App (DMOVA) in a university context in Vietnam. Data were collected from surveys, semi-structured interviews, observations and speaking tests.

This chapter presents the findings from an analysis of the collected data. SPSS was used to calculate Cronbach's alpha reliability coefficients and highlight correlations between the live and digital marking results. Coding and analysis of the responses to open questions in the surveys and teacher interviews, as well as the teacher and student observations, were undertaken with NVivo 12, a qualitative data analysis software. The findings are presented according to the data collection methods that included surveys, observations, teacher interviews and the test results database.

Survey Data

By the end of the survey period, data were collected from 60 students (N(S2) = 60) and 18 teachers (N(T2) = 18). The student survey was conducted after videos of their speaking performances were returned to them. The Cronbach's alpha reliability coefficient for internal consistency of the 80-item Likert-scale student survey was 0.98, which could be considered excellent reliability given the range proposed by George (2011). The teacher survey was administered after they had finished marking the student performances. The Cronbach's alpha reliability coefficient for the 82-item scale was 0.97, indicating high internal consistency and reliability of the measuring instruments.

Teacher Survey

Demographic Information

Eighteen teachers participated in Phase 2 of the research (N(T2) = 18). Fourteen teachers were female and four were male. Half were aged between 26 and 35 and seven were between 36 and 45. Only two teachers were under 26 and over 46 respectively. Thus, the age range was between 26 and 45.

Table 5.1

Age Groups of Teacher Participants

Age group	Number represented in population (N(T2) =18)
≤ 25	1
26 - 35	9
36 - 45	7
≥ 46	1

As shown in Table 5.2, the majority of teachers had several years' experience teaching EFL. A large number had been teaching English for six to ten years, and nearly half, for over 10 years. The numbers were distributed quite evenly for years of teaching English. The same number of teachers (4) had been teaching English for less than 5 years as from 11 to 15 years and over.

Table 5.2

Teachers' Years of Teaching English

Years of teaching English	Number of the teachers (N(T2) =18)
0 – 5 years	4
6 – 10 years	6
11 – 15 years	4
Over 15 years	4

In summary, the teacher participants had similar characteristics regarding age and teaching experience. Most were between 26 and 45 years old and had been teaching English for 6 to 15 years. The relatively young age of most teachers was a reflection of the recent establishment of FPT University in 2006.

Teacher Experience

Teachers (N(T2) =18) were asked about their experience and familiarity with computer-assisted EFL tests. In this study, experience was understood to be teachers' use of these tests and familiarity was defined as frequent use. Fifteen teachers reported using, adapting, designing and delivering computer-assisted English tests. The same number replied that they were interested in and familiar with using, adapting, designing and delivering computer-assisted English tests. Sixteen teachers agreed that computer-assisted tests outnumbered paper-based tests at the university. The results showed that the majority of English teachers at FPT university were experienced and familiar with using ICT in EFL assessment.

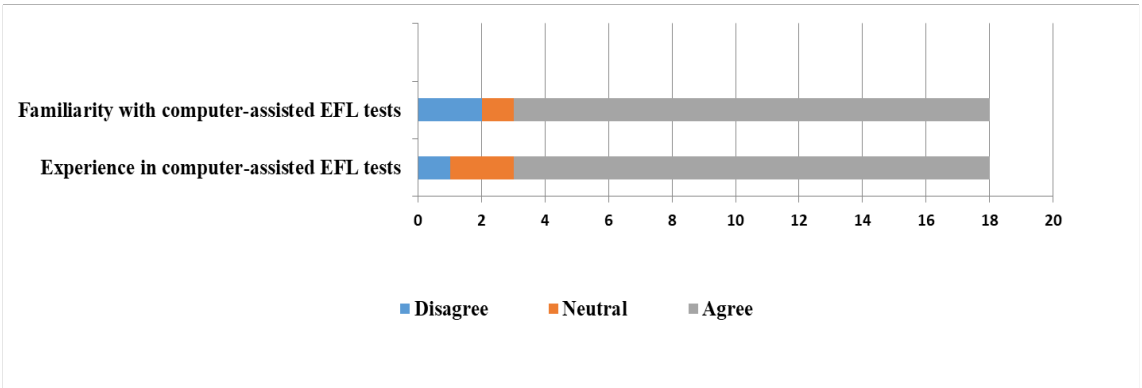


Figure 5.1 Teacher Experience with Computer-Assisted EFL Tests.

As shown in Figure 5.1, there was a small number of teachers who did not have any experience with computer-assisted English tests. There was also a small number that provided neutral responses, possibly due to a lack of experience with computer-assisted EFL tests.

Computer-Assisted Speaking Tests

Figure 5.2 shows teachers' use of computer-assisted tests across the different language skills. Seventeen teachers claimed that they used, adapted, designed and delivered computer-assisted reading tests. A large number agreed that they used computer-assisted tests to check students' competency in grammar (16), vocabulary (14), and listening (13).

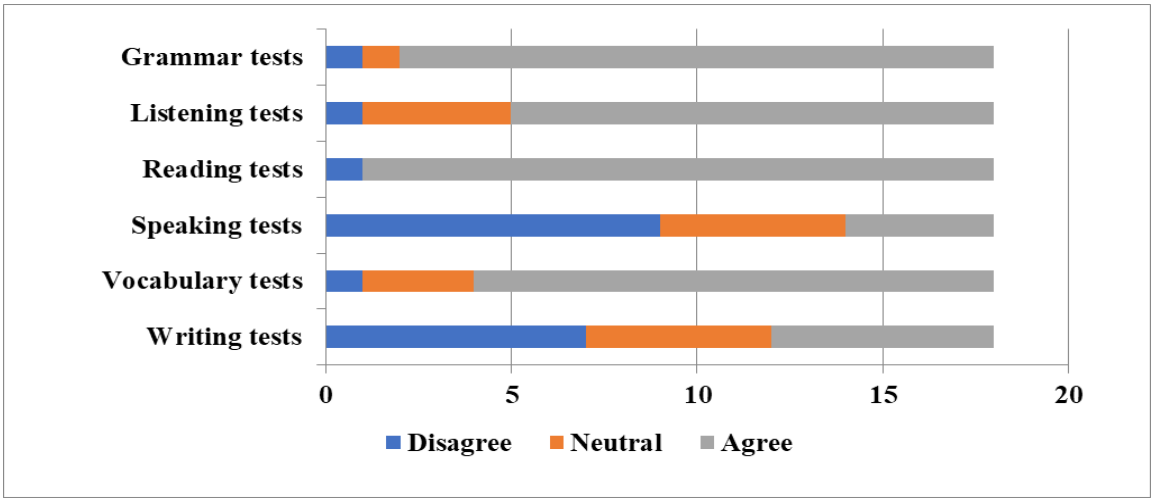


Figure 5.2 Teachers' Use of Computer-Assisted EFL Tests.

A minority of teachers (6) said they used computer-assisted tests to check their students' writing skills. Only four used, adapted, designed and delivered computer-assisted tests to check students' speaking skills. As shown in Figure 5.2, out of the six types of skills,

speaking skills were the least tested this way. The data also suggested a higher frequency of computer-assisted tests for assessing receptive skills (reading and listening) than productive skills (writing and speaking).

Although few teachers used computer-assisted English speaking tests, they seemed to integrate ICT more into other teaching activities. The survey showed that a large number of teachers recorded videos of their student speaking performances for assessment (11), assigned students tasks of videoing their presentations and practicing at home (13) and used them for assessment purposes (14). The results also showed that ICT was not popular for assessing speaking and English teachers had acquired some experience with it elsewhere.

Teacher Beliefs about DMOVA

After digitally marking the student speaking performances, the teachers' perceptions and experience with DMOVA were explored via a survey.

Capturing Speaking Performance

Most teachers (14) agreed that the sound and image quality of the videos were more than adequate for marking. One teacher claimed enthusiastically that these factors enhanced the accuracy of assessments. Fifteen teachers agreed that the videos were a true representation of student performances. Three teachers complained about the sound quality of some videos.

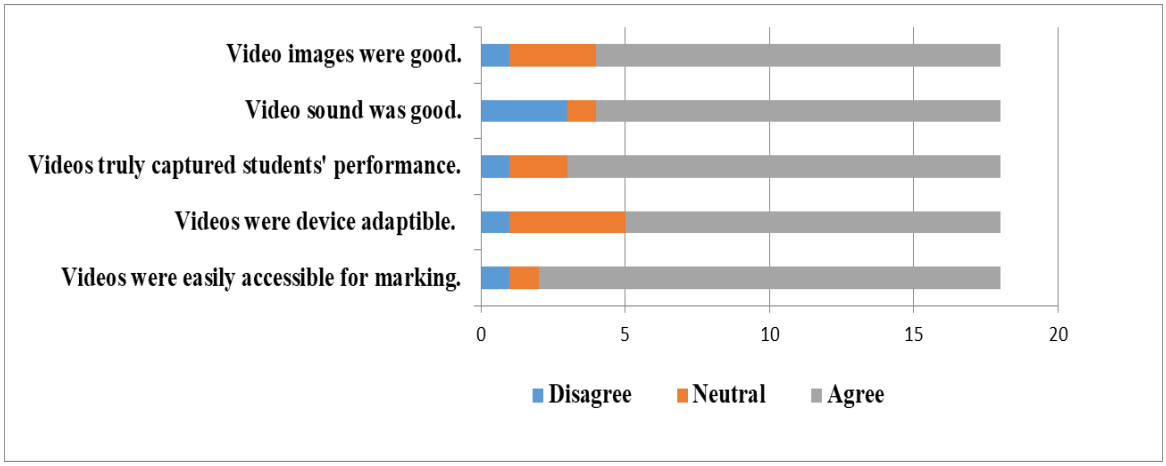


Figure 5.3 Quality of the Videos.

One teacher commented that the iPad on which the videos were recorded did not have a good voice recorder, so the sound was difficult for her to hear and mark (Q12 - Responses). She added that better quality equipment may have to be provided to resolve the audiovisual issues (Q13 - Responses).

Another teacher noted the individual performances had better sound quality and less interference than the group performances. As a result, she found the individual task videos easier to listen to (Q14 - Responses). Another recommended using a special acoustic room for speaking tests with video recordings (Q20 - Responses).

Thirteen teachers agreed that digital representation was compatible with numerous digital devices, including iPads, laptops, smartphones, and iMacs. Sixteen agreed that easy access to the videos via an internet browser gave them more flexibility to mark at a time and place of their convenience. Easy accessibility was also credited with enabling multiple reviews and checking (Q12 - Responses).

Some teachers had doubts about the effectiveness of assessing English speaking skills from digital representations. One raised concerns about the cost of equipment (Q13 - Responses). Forgetting to press the record button was also mentioned by some (3). Another teacher pointed out that failure to record was due to human error on the part of invigilators and called them absent-minded mistakes (Q13 - Responses).

Transparency of Assessment

Fourteen teachers believed that DMOVA was an effective way of evaluating student speaking performances, and fifteen agreed that it highlighted previously unnoticed strengths and weaknesses.

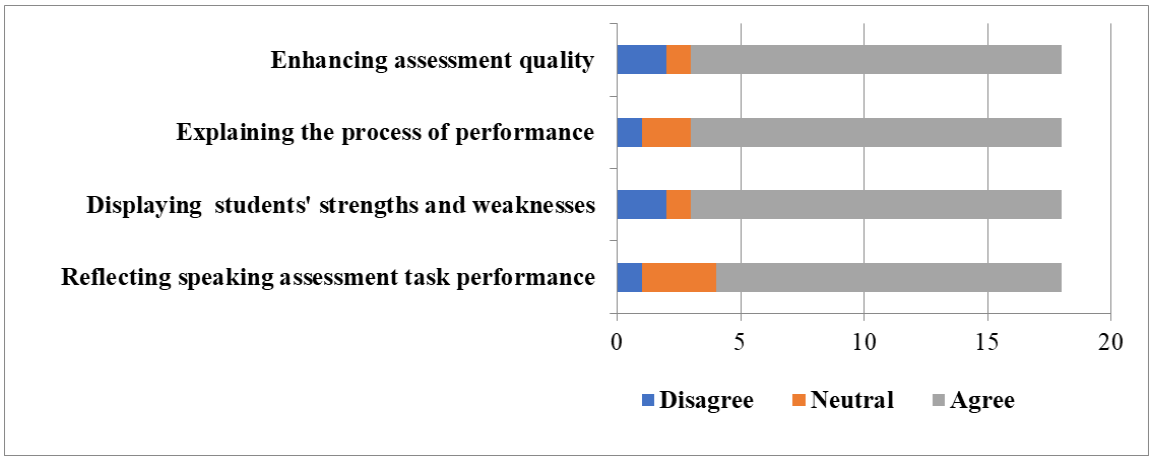


Figure 5.4 Benefits of DMOVA for Speaking Assessments.

They concurred that DMOVA was useful for describing the student performances, i.e., how they dealt with the test questions, how they interacted with one another in group tasks, and how they started and concluded their talks. Insofar as these aspects were concerned, they believed the digital method was on task to enhance assessment quality.

Teachers commented on the convenience and flexibility of DMOVA: “time-saving and highly efficient in marking without reducing the quality of assessment” (Q12 - Responses). They believed it “enhanced fairness” and provided “precise results”, “easy review”, “good visual and sound quality, high level of accuracy in assessing students’ English competence” (Q12 - Responses).

Seventeen teachers reported that DMOVA effectively supported speaking assessments. Sixteen agreed it was good for recording student performances for practice and assessment. A large number (16) were optimistic about the reliability and feasibility of the new testing method. Most (16) were interested in using digital representation for speaking assessments in the future.

The majority of teachers testified that DMOVA was effective for both individual and group assessment tasks. Three teachers found it more suitable for group tasks because “teachers can give more exact marking” by comparing and contrasting individuals in the groups and observing their interactions (Q14 - Responses). Four others claimed it was more effective with individual tasks: “It was easier to focus on each of the students than a group of students talking” (Q14 - Responses), stating that the individual recordings were free from interference by other group members and easier to listen to. Overall, the teachers believed that the digital representation enhanced individual assessment of student speaking skills.

Performance Backup

Sixteen teachers positively endorsed the benefits of DMOVA in terms of its usefulness for backup purposes and liked the flexibility of reviewing the videos at their convenience. The same number cited the advantages of providing evidence of student speaking performances and exam attendance. Seventeen teachers claimed that digital representation served as records of student performances in the same way as other EFL skills assessments, emphasising its disparate standing and lack of attention.

Ten teachers acknowledged the significant benefits of backing up digital performances. “Backup for future review”, “keep recordings of students’ performance”, “backup and teachers can check the students’ performance again”, “recheck”, “remark”, and “review” were all frequently mentioned in response to the open survey questions (Q12 - Responses).

Motivation

Sixteen teachers observed their students were better prepared for their speaking tests when they knew their performance was going to be videoed. Fifteen witnessed improvements in their students' speaking, such as using gestures, correct posture, eye contact, and facial expressions, as well as fluency and richer content. According to the teachers, students were motivated to perform better when they were videoed; sixteen agreed that digital assessment of speaking skills had the potential to boost student learning and teacher motivation.

Although relatively positive about the benefits of DMOVA, a small number of teachers were doubtful. They were concerned about a possible lack of student-teacher interaction and that they "could not give instant feedback to students". They also worried that students might not be confident in front of the camera and that technical problems could disrupt testing (Q13 - Responses).

Management and Adaptibility

Eleven teachers commented on the ease of managing the technologies and the test at the same time. Twelve confidently concluded that one invigilator could manage the technologies and organise the test without assistance. Ten teachers were of the view that DMOVA eliminated the need to employ English test invigilators and solve the current shortage of English invigilators every semester. The majority of teachers (13) were also optimistic that the available facilities at the university adequately supported digital assessment.

Most teachers were positive about the compatibility of DMOVA with the existing technologies at the university and its capacity to support management. However, six teachers had doubts about the authenticity of speaking tests delivered by an invigilator who was not an English teacher. They argued that EFL teachers were still necessary to ensure the test wasn't cancelled due to technical problems, in which case they could take over and complete it themselves.

Overall, the majority of teachers (15) believed that digital representation was effective for assessing EFL speaking skills; only three were doubtful. In comparing DMOVA with the current method, twelve teachers considered the digital method a better option. One third of the teachers surveyed (6) gave neutral responses.

Flexibility

Figure 5.5 shows all surveyed teachers (18) agreed that DMOVA gave them flexibility to review student performances and do the marking when it was convenient. “Teachers can check the students' performance again” and “can mark anywhere anytime” (Q12 - Responses). Question 12 of the survey recorded ten responses to “benefit of backup for later review”, and six other responses regarding time saving and flexibility for marking.

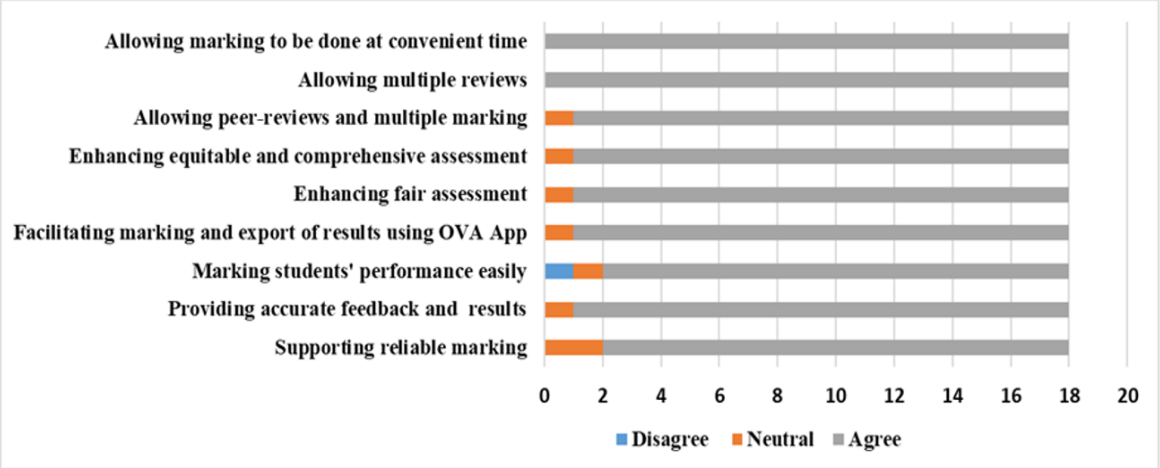


Figure 5.5 Impact of DMOVA on Speaking Assessments.

Seventeen teachers reported that the new testing method made a real difference because they could watch and listen to the videos multiple times. This allowed them to provide students with more detailed feedback and more accurate results (Q12 - Responses). The same number of teachers (17) claimed the OVA App facilitated their marking and they could easily export the results. The majority (16) found the digital representation easy to mark.

Analytical Marking Method

Figure 5.6 shows an increase in analytical marking for DMOVA assessments, indicating a difference in marking methods between the current and digital modes. In the current method, teachers commonly used a combination of analytical and holistic marking, with some (6) using only analytical marking. None of the teachers reported marking holistically when invigilating current speaking tests.

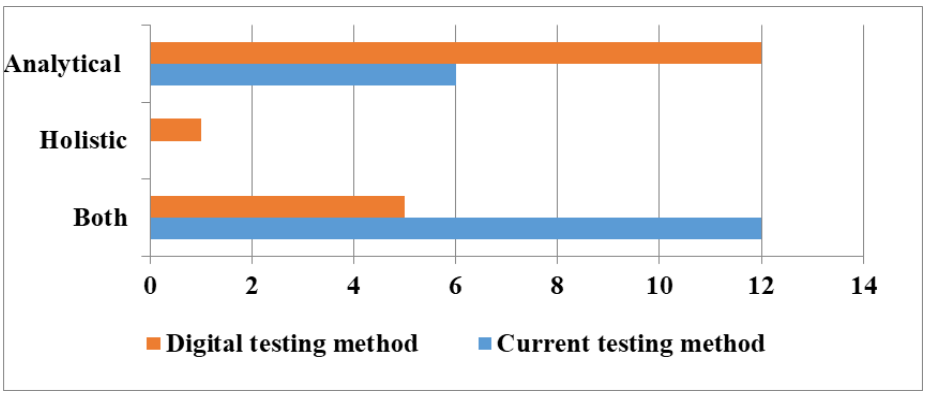


Figure 5.6 Teacher Marking Methods.

Twelve teachers claimed they mainly used the analytical method to mark the digital performances, in close alignment with the marking key. One marked holistically and five others used a combination of the two methods. There was a distinct increase in the use of analytical marking with digital assessment.

Teachers proposed recommendations for the marking key, which was adapted from the existing one at FPT University. Most suggested the inclusion of additional categories and benchmarks. One teacher said: "The marking criteria for the individual tasks should be more detailed to cover the range of speaking ability". Another teacher asked about using half marks (e.g., 0.5) for grading (Q17 - Responses).

Peer Review and Multi Marking

Seventeen teachers were enthusiastic about DMOVA's capacity to allow peer-review and multi-marking of student performances. The same number also agreed that it enhanced fair marking compared to the current method. Moreover, they believed that DMOVA helped them assess speaking skills more equitably and comprehensively. The teachers pointed out that, thanks to the advantage of being able to replay videos multiple times, it would be difficult to miss important aspects of student performances, common mistakes and individual weaknesses. Most believed that DMOVA facilitated providing students with more accurate results.

Marking Reliability

Sixteen teachers expressed the view that digital marking was more reliable for speaking assessment than the traditional paper-and-pencil method. Two teachers were neutral and none disagreed. They found it easy to mark individual assessments, identify individuals in the group tasks, and had no difficulties marking group tasks and entering feedback into the OVA App. One teacher commented that "it was easier to focus on each of the students than a group of students talking" (Q14 - Responses). Another teacher reported

wasting time marking the group tasks because she had to replay the video four times, one for each student in the group (Q13 - Responses). A further teacher admitted that she sometimes felt the urge to fast-forward the videos and speed up her marking at the risk of missing important aspects of the performance. She was also concerned that teachers could not provide instant feedback with digital assessment as they could with direct interviews (Q13 - Responses).

Impact on testing, teaching and learning

The fairness and accuracy offered by digital marking appeared to have had an overall positive impact on English teaching, learning and testing. All the teachers (18) agreed that the ability to save their feedback in the DMOVA results database and send it to their students was a distinct advantage. Students would be able to clearly identify aspects of the language they needed to improve for better results in future speaking tests.

Sixteen teachers stated that the process of marking with DMOVA helped them understand their own shortcomings and see how they could improve. One teacher focused more on the performance and marked with more detail using the marking key. Another teacher claimed that digital marking gave her more time to consider each student's strengths and weaknesses and compare results.

Benefits for Testing and Teaching

Figure 5.7 shows nearly all the teachers (17/18) agreed that DMOVA would be valuable for reviewing student performance after exams. They also recognised its potential for assigning homework to students and backing up their performances.

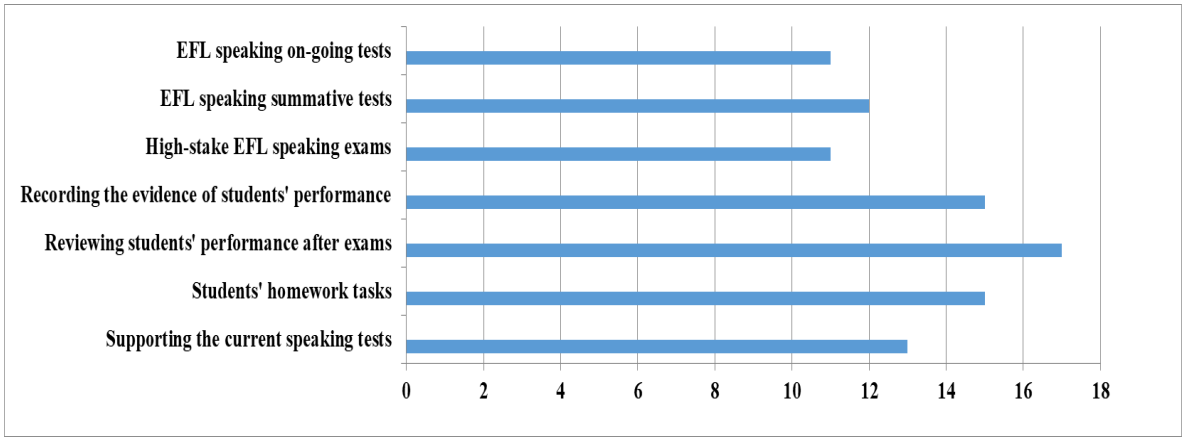


Figure 5.7 Perceived Effectiveness of DMOVA.

More than half the teachers proposed that digital marking be used to supplement the current method. They considered it an effective tool for summative, ongoing speaking tests and high-stakes exams. One teacher suggested using DMOVA to observe teacher assessment practices (Q21- Responses).

Teacher Preferences

Figure 5.8 shows that teachers preferred the new marking method in relation to DMOVA's backup, flexibility, reliability and validity features. However, in relation to economical features, pedagogical effects, ease of practice and effectiveness, they preferred the current method.

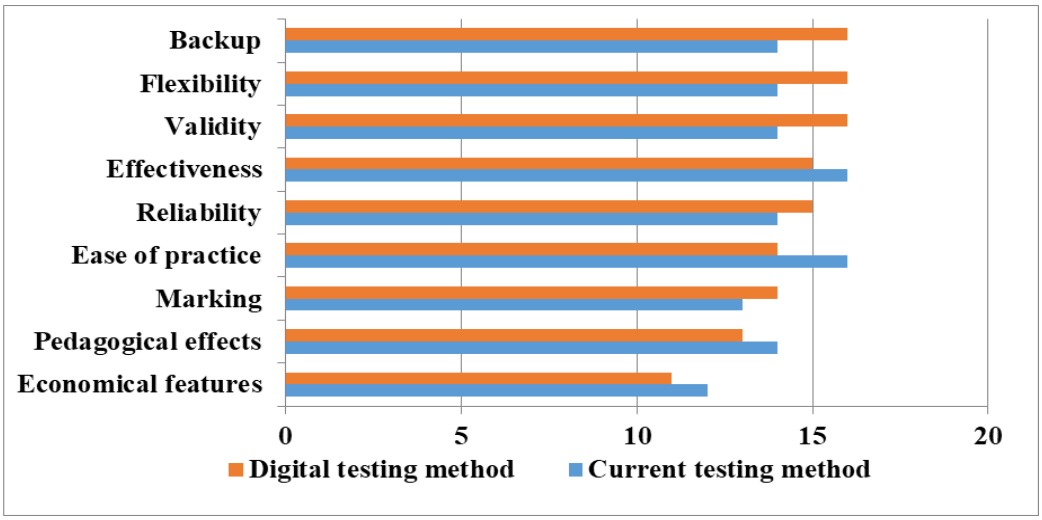


Figure 5.8 Teacher Perceptions of the Current and Digital Testing Methods.

Teachers liked that digital assessments allowed them to review student performances, recheck results and make comments. They agreed that DMOVA facilitated efficient marking "without reducing the quality of assessment" and gave them more time to mark thoroughly and compare students' speaking competencies. They also responded positively to the convenience of marking anywhere, anytime (Q12 - Responses).

Some teachers mentioned that students' fear of detection on video may deter cheating (Q12 - Responses). Although the survey results showed they were happier with the reliability, validity and flexibility of the digital testing method, some teachers were concerned about the lack of student-teacher interaction (Q13 - Responses). This was also the reason for their low satisfaction with the pedagogical impacts of DMOVA.

Most responses related to backup advantages. The largest number of respondents praised the ability of DMOVA to record student performances as backup of student performances for assessment and future review. They reckoned that keeping recordings

of speaking tests would level the playing field with assessments of other language skills (Q 12- Responses).

The teachers who were doubtful said: “It takes time to set up and probably needs team support. It’s difficult for an invigilator to do it alone”. Their concerns ranged from: “expensive supporting devices” to: “the devices that we use to record may run out of batteries and have technical problems” (Q13 - Responses). Teachers recommended checking the devices in advance of tests to ensure they were functioning properly. One described the dependence of digital marking on technical equipment, batteries and the internet as a deterrent. Another was worried about test disruptions and wasting time if the equipment failed. Overall, teachers expressed a lower level of satisfaction with the economical features of the digital testing method. Despite these issues, they noted that the digital method offered convenience and saved time and human resources. It also ensured fairness and reliability and they could mark at convenient times and locations (Q12 - Responses). One teacher expressed concern about the availability of team support and extended setup times (Q13 - Responses).

Sixteen teachers concurred that the digital testing method smoothed the process of managing tests and test results. They could retrieve the results after the test and remark if necessary. Fifteen teachers endorsed the practicality and feasibility of DMOVA in the context of FPT University.

Some teachers raised the issue of students’ discomfort in front of the camera, reporting that they lacked confidence when they were videoed. They felt shy and stressed and therefore did not perform at their best (Q13 - Responses). One teacher observed some students displaying confidence in front of the camera and enjoying their “freedom” (Q12 - Responses).

Teachers proposed adding technical features to the OVA App for marking pronunciation (Q17 - Responses). The OVA App “should also support offline. Teachers may also be able to download the videos and assess offline and may sync or upload the results later.” In this way, teachers “do not have to be completely dependent on the internet connection” (Q20 - Responses).

Summary

In summary, analysis of the teacher surveys highlighted the following findings:

- The majority of teachers indicated they were experienced and familiar with computer-assisted EFL tests,

- Of the six types of English skills, speaking was the least assessed by means of computers,
- DMOVA was considered effective for assessing speaking skills. The digital representation captured student speaking performances, enhanced assessment quality, supported backup, motivated teachers and students, assisted management, and was compatible with the existing technologies at the university,
- DMOVA was found to facilitate marking, enhance assessment quality and have a positive impact on English teaching and learning,
- DMOVA provided perceived benefits for different testing and teaching activities,
- Teachers expressed positive attitudes towards the digital testing method.

The findings of the teacher survey in Phase 2 triangulated with the findings of the teacher survey in Phase 1 as follows:

- The majority of teachers indicated they were experienced and familiar with computer-assisted EFL tests,
- They expressed a preference for computer-assisted EFL tests,
- They had little experience and practice with adapting, designing and delivering computer-assisted EFL speaking tests in their English classrooms,
- They expressed positive attitudes towards computer-assisted EFL speaking tests.

The findings of the teacher data collected in Phase 2 confirmed the findings of the teacher survey in Phase 1. Further findings are presented in the analysis of the observation data.

Student Survey

Demographic Information

The demographic characteristics varied for the 60 student respondents to the survey (N(S2) = 60) as shown in the tables and graphs below for the purpose of comparison and contrast. The students were in semester two of their first year at university. Their age distribution is shown in Table 5.3. A large majority (93.4%) were between the ages of 19 and 20, with a small percentage 21 and older. The oldest student was 23 at the time of completing the survey. In general, therefore, students were roughly the same age.

Table 5.3

Student Age Groups

Age group	Percentage in the population (N(S2) = 60)
19 - 20	93.4 % (56)
21 - 22	3.3% (2)
≥ 23	.3% (2)

Their gender composition was 87% male, 11% female and 2% (one student) of unidentified gender. FPT University was a technical school, and according to its gender statistics, male students usually outnumbered females. The above gender distribution is typical of technical university students in Vietnam (Dang, 2016). For example, according to the statistics for Ho Chi Minh National University (2016), more than 80% of students at the Polytechnics University and Information Technology University were male (Dang, 2016).

Most of the student respondents (67%) had been learning English for between seven and ten years. Eight percent had been learning English for more than 10 years. Table 5.4 indicates a small number of students had learnt English for less than six years, while the majority had been learning English for seven years or more.

Table 5.4

Years of Learning English

Years of learning English	Percentage represented in population (N(S2) = 60)
0 - 3 years	11 (18%)
4 – 6 years	4 (6.7%)
7 – 10 years	40 (67%)
>= 10 years	5 (8.3%)

Student Familiarity with Computer-Assisted Tests

Table 5.5 presents data on student experiences with taking computer-assisted tests in all their university subjects. Approximately 90% had taken such tests before. More than 75% indicated they were used to taking computer-assisted tests. Nearly 65% of students expressed a liking for computer-assisted tests, while 26.7% were neutral. A total of 88.3% of students reported that computer-assisted tests were popular at their university and far outnumbered the paper-and-pencil test method.

Table 5.5

Computer-Assisted Tests at FPT University

(N(S2) = 60)	**Disagree**	**Neutral**	**Agree**
Experience with Computer-assisted tests	5 (8.3%)	1 (1.7%)	54 (90%)
Familiarity with Computer-assisted tests	7 (12%)	8 (13%)	45 (75%)
Interest in Computer-assisted tests	5 (8.3%)	16 (26.7%)	39 (65%)
The frequency of Computer-assisted tests	2 (3.3%)	5 (8.3%)	53 (88.4%)

Student Experience with Computer-Assisted EFL tests

The results showed that 91.7% of the student participants had taken computer-assisted EFL tests at university. Seventy-seven percent were accustomed to taking these types of language tests and 65% expressed an interest in taking English tests on computers, while 25% were neutral and a small minority did not like taking English tests on computers. More than 83% said that computer-assisted EFL tests were more popular than paper-and-pencil assessments (see Table 5.6).

Table 5.6

Computer-Assisted EFL Tests at FPT University

(N(S2) = 60)	**Disagree**	**Neutral**	**Agree**
Experience with Computer-assisted EFL tests	4 (6.6%)	1 (1.7%)	55 (91.7%)
Familiarity with Computer-assisted EFL tests	8 (13%)	6 (10%)	46 (77%)
Interest in Computer-assisted EFL tests	6 (10%)	15 (25%)	39 (65%)
The frequency of Computer-assisted EFL tests	6 (10%)	4 (6.7%)	50 (83.3%)

Neutral and disagree responses to this item could be explained by the fact that, at the time of the research, there was a small number of international students newly enrolled in the English intermediate level and a few new students had arrived from other universities who may not have experienced computer-assisted tests (Teacher 1, Interview, 2018).

Figure 5.9 shows that computer-assisted tests were popular at FPT University and were used in subjects other than English. Students expressed an interest in computer-assisted tests in all their subjects and were confident of their abilities to undertake them successfully.

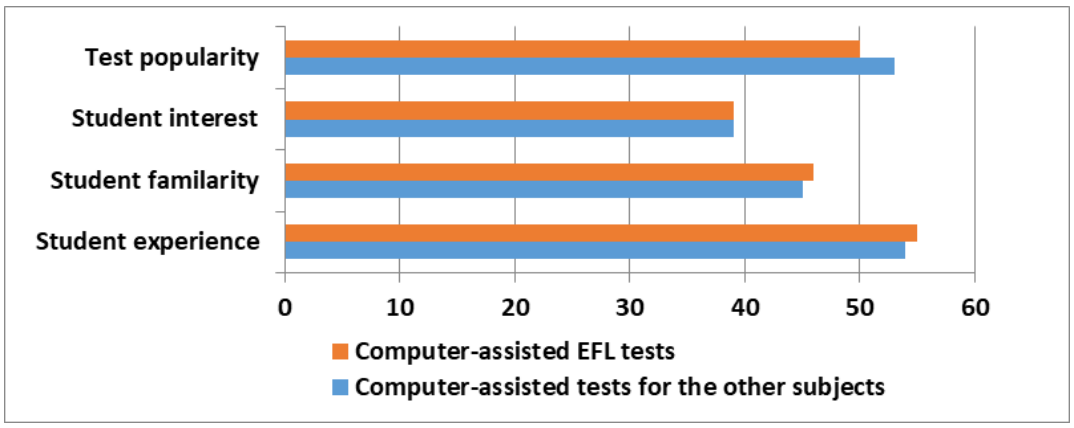


Figure 5.9 Computer-Assisted Tests at FPT University.

Computer-Assisted Tests for EFL Speaking and Writing

Figure 5.10 shows that ICT was integrated in all English skills testing at the time of the research, including reading, listening, writing, speaking, grammar and vocabulary. However, the frequency of use was different for each skill. The majority of students regularly sat digital English grammar (87%) and vocabulary tests (82%), and many were also familiar with computer-assisted listening and reading tests. Writing and speaking skills were the least tested in this way. Almost 42% of students had never undertaken English speaking tests with ICT integration and 15% were not sure whether they had. Forty-seven percent reported that computer-assisted English writing tests were completely new to them.

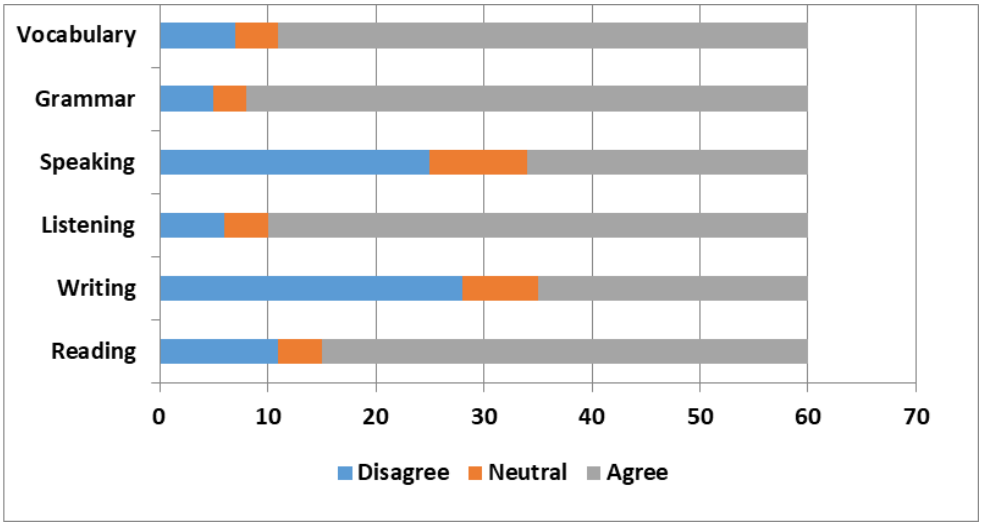


Figure 5.10 Frequency of use of Computer-Assisted EFL Tests.

Although the data showed that few students had taken computer-assisted English speaking tests, further investigation revealed that many of them had recorded videos of their English speaking performances for assessment (63%) and practice (65%) (see

Figure 5.11). Therefore, video recordings of their English speaking performance may not have been completely new to them, and they may have come to the test trial with experience and confidence to pose in front of the camera.

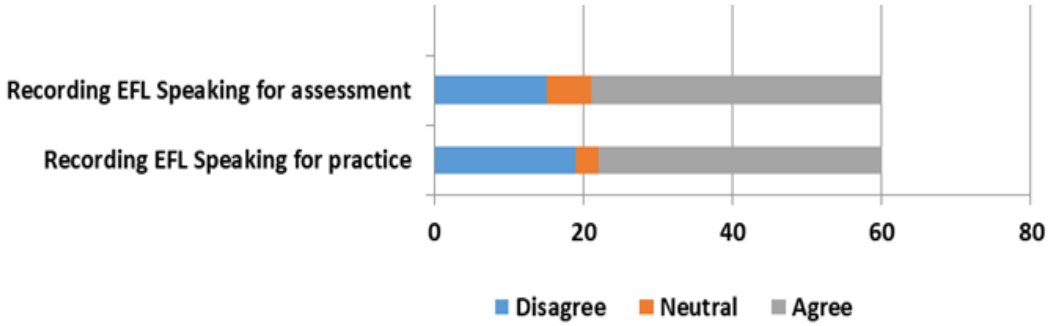


Figure 5.11 Video Recordings of English Speaking Performances.

Student Beliefs about the Benefits of DMOVA

Benefits for EFL Speaking Assessment

Eighty seven percent of students found DMOVA an effective way to authentically capture their speaking performances. They commented on the high sound and resolution quality (Q13 -Student responses) of the videos and made improvements by adjusting the position of the camera to best capture their performance (Q14 - Student responses).

Over 80% of students viewed DMOVA as an effective way of explaining the process of performance and for supporting marking and review. Ninety two percent agreed that digital representation provided a record of performance, similar to the other English language skills of reading, writing, and listening. Over 45% of students talked about the benefits of digital representation for backing up test performance and allowing teachers to remark and review. The most common responses to the open survey questions were: “keep the recording of students’ performance”, “backup”, “review”, and “remark”. Students also anticipated being able to check their results and refer to teachers’ feedback multiple times after taking the test. One student remarked: “We can see the results many times later” (Q13 - Student responses).

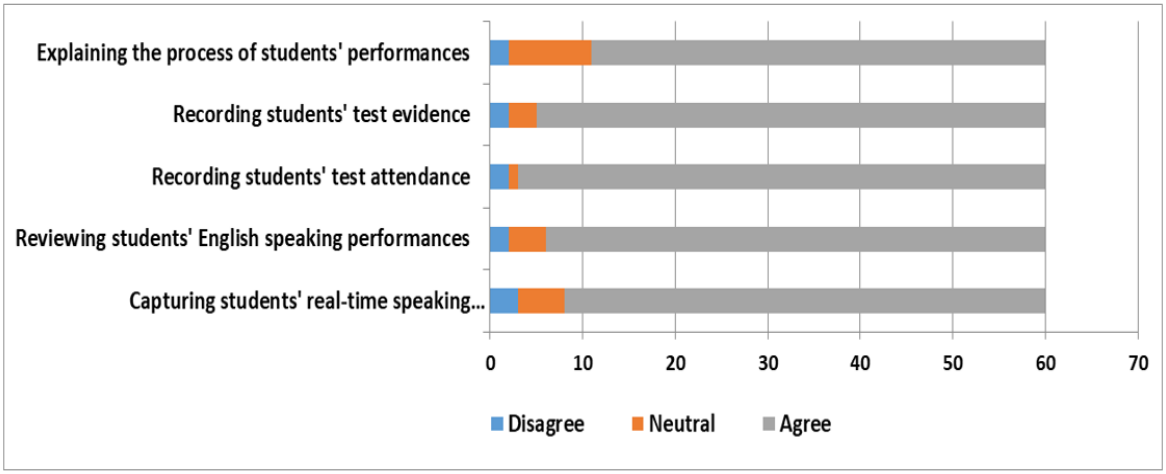


Figure 5.12 Student Perceptions of the Benefits of DMOVA.

Most students (95%) agreed that the digital records would serve as evidence of their exam attendance and performance. Ninety percent of them also affirmed the advantages of being able to review their own records and for markers to review their results.

Benefits for Student EFL Speaking Skills

Ninety three percent of students reported that the videos helped them recognise their strengths and weaknesses by watching themselves perform. One student wrote: “I can watch and re-watch my video multiple times to recognise my weaknesses and my common mistakes in my speaking, then I will avoid them later”. Another student wrote: “I can watch the video many times and I myself will know my level of English speaking skills” (Q13 - Student responses). Students were also of the view that watching the videos would enable teachers to see the results of their practice and efforts to improve their speaking skills.

Seventy eight percent of students expected the digital representation would encourage their learning of speaking skills, better prepare them for speaking tests and focus more on their execution, not merely on the content of their interaction. The knowledge that they were being recorded and could be marked by several teachers was the incentive they needed to put their best foot forward. One student claimed that after watching his own video and receiving feedback from the teachers he “could fix my mistakes in speaking English” (Q13 - Student responses). Students also perceived that the new testing method would help prevent cheating and therefore enhance fairness.

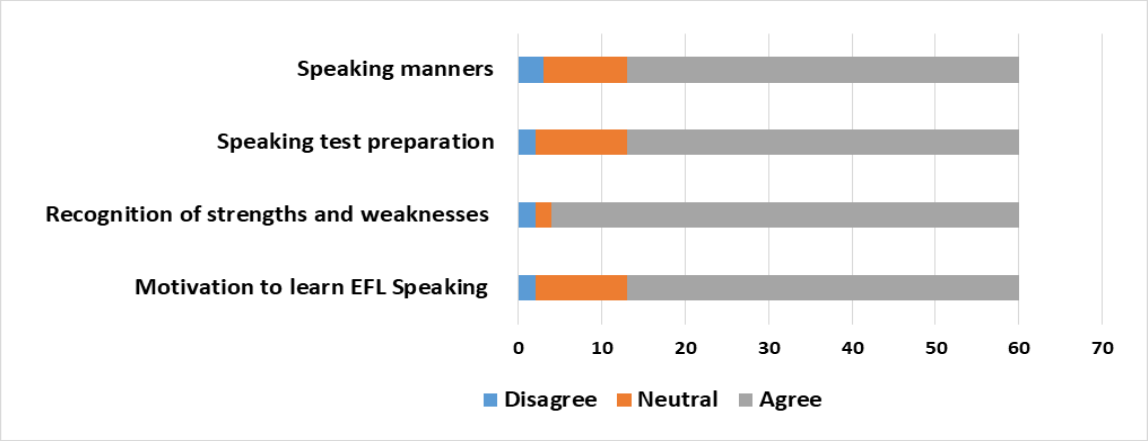


Figure 5.13 Benefits of Digital Representation.

Seventy two percent of students agreed that DMOVA enhanced their assessment results, thanks to the positive impact of this method on motivating them to learn and improve their performances. One student explained that, given digital representation generated accurate marking, this indirectly motivated students to improve their speaking skills (Q13 - Student responses).

Overall, approximately 80% of the student cohort believed that digital representation was an effective method for English speaking assessment. More than 90% agreed it was more accurate and effective than the paper-and-pencil method, as well as more objective and reliable. Some commented that the new testing method was fast, easy to use, and facilitated management of their performance and test results (Q13 - Student responses).

Perceptions of Reliability and Feasibility of DMOVA

Seventy two percent of students made positive comments about the reliability and feasibility of digital representation. In response to the open questions they stated that the digital testing method was “reliable” (9 responses), “objective” (5 responses), “fair” (14 responses), “accurate” (11 responses), and “convenient” in terms of easy accessibility (13 responses). Three quarters of the students believed that DMOVA was a more reliable form of assessment than the current method, and 65% indicated they enjoyed using the digital format.

Based on the survey results, many students did not perceive performing in front of the camera a big challenge. Thirty two percent displayed their confidence in the test room. Fifty percent reported feeling okay about being videoed and 45% replied that they liked having their performance recorded. One student explained that he gradually got used to standing in front of the camera. He found the new testing method ensured fairness and produced high quality assessment results (Q13 - Student responses).

Figure 5.14 shows the perceptions of students towards different aspects of the digital presentation process. Videoing the test gained the highest satisfaction rate, with 71.7% of students judging it positively. The technologies used for the tests also received a high rate of satisfaction (70%). Sixty percent of students agreed that both individual and group tasks were satisfactorily facilitated by the digital method. Over 70% were positive about the test room setup. The waiting time before tests and the time needed to finish the test satisfied 65% of the students.

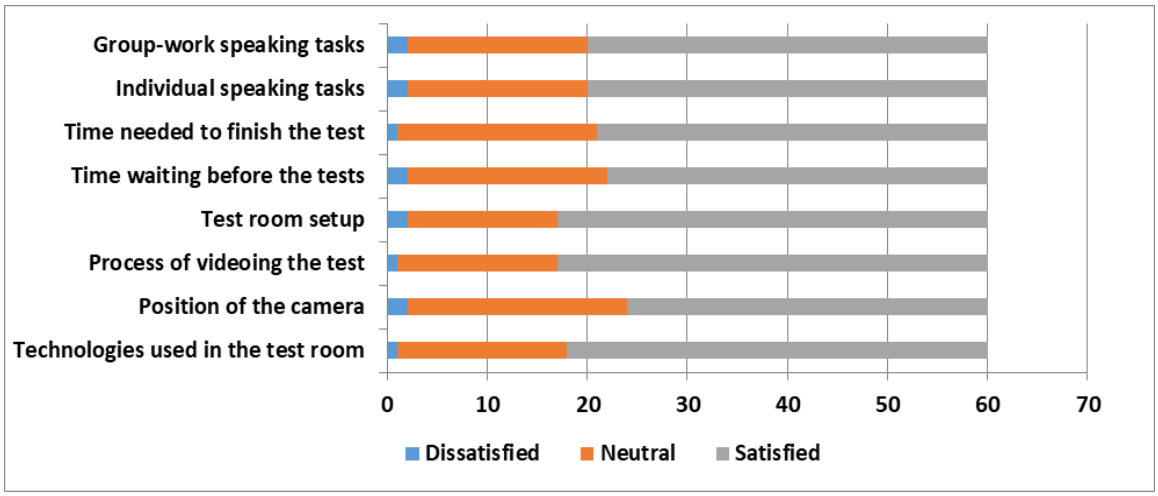


Figure 5.14 Student Perceptions of Digital Test Setup.

The large number of neutral responses was noteworthy (see Figure 5.14). The position of the camera in the test room received the most responses (37%). Many students (33%) did not show clearly whether they were satisfied or dissatisfied with the waiting time and the time needed to complete the test. It could be that more experience will cement their opinions of the digital testing system. It is also possible that the students who returned neutral responses were critical of the new testing system and provided suggestions on how to improve testing procedures in the open response section of the survey. Figure 5.14 indicates that the overall number of students who were dissatisfied with the digital testing procedure was under 4%.

After experiencing the digital testing method, a little over a third (35%) of students said they were nervous and shy about being video recorded. Nearly a quarter said they did not feel good about being videoed. When asked what they did not like about digital representation, 30% cited feeling stressed and lacking in confidence in front of the camera because this way of testing was unfamiliar to them.

Some students expressed concerns about the feasibility of the new testing method in terms of data security and economy. One was concerned about technical problems that

might arise during assessments, such as recording failure, and lead to test delays and cancellations (Q14 - Student responses).

Perceptions of Equitability and Comprehensive Assessment

Question 9 of the survey related to how the speaking performances would actually be assessed. Ninety two percent of students agreed that DMOVA was very different from the current method, in that it allowed markers to watch and listen to student performances multiple times. Therefore, they assumed, markers would provide more detailed feedback and more accurate results.

Ninety percent of students believed that the digital method encouraged markers to assess speaking skills more equitably and comprehensively because DMOVA afforded them more time to do their marking compared to the live marking method. Eighty three percent of students considered the new testing method more reliable. The digital representations meant that markers could assess the performance as a completed work rather than a live ongoing performance.

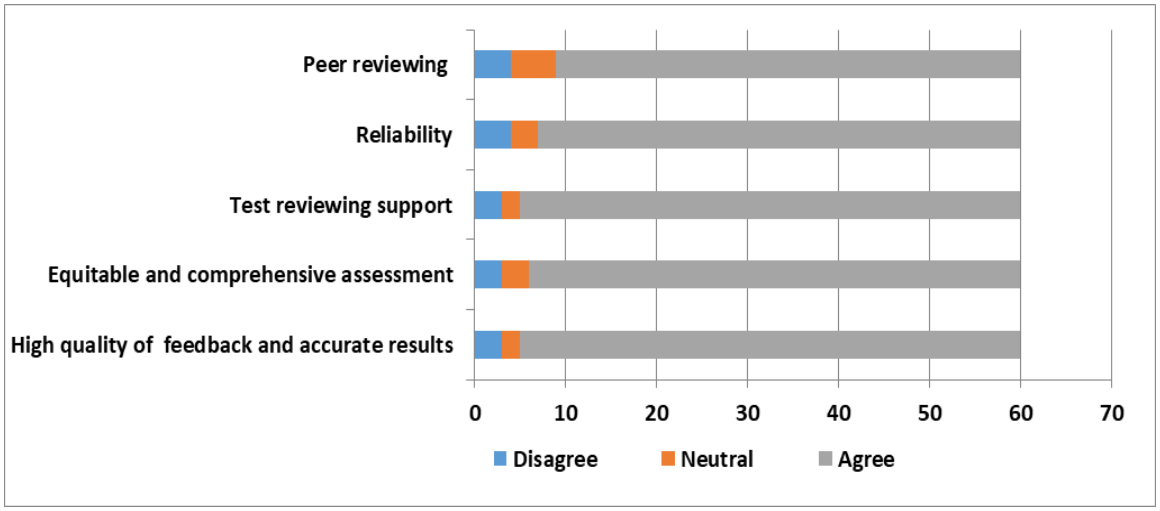


Figure 5.15 Student Perceptions of DMOVA.

A large number of students (92%) acknowledged the benefit of recording their performances for later review. The current testing method at FPT University did not record student speaking performances, which made it impossible for markers to review their work later. Eighty eight percent of students liked the DMOVA feature for recording markers' feedback, as this not only helped them understand their strengths and weaknesses, but also inspired them to improve their performances. A large majority of students (85%) were keen to share their performance videos with peers and other teachers for additional feedback and comments, in recognition of the opportunities for learning from their own and others' mistakes.

Overall, the students surveyed were positive about the quality of DMOVA. They were most positive about the benefits related to recording performances for later review, the high level of accuracy, and quality of the feedback from markers.

Satisfaction with DMOVA

Although the students were happy with the current testing method for speaking, they were even happier with the digital method. The data indicated that the students were less satisfied with the current English speaking test management, organisation, and distribution of results than those same aspects of the new digital method. Eighty three percent of students were satisfied with DMOVA, while 68% were happy with the current testing method. “Easy to manage”, “easy to share videos and results”, “I can watch my own performance”, “professional”, “modern”, and “innovative” were some of the student responses to questions about test management, organisation and distribution.

The survey data showed a large gap in student satisfaction with the backup capability of the digital method at 80% and the current method at 62%. Almost 40% of student responses to the open questions mentioned the backup advantages of the digital method with responses like “recording students’ performance”, “backup”, “allowing reviewing”, and “record and confirm the authenticity of students’ performance” (Q13 - Student responses).

There was also a higher level of satisfaction with the marking process of the digital assessment method. Seventy eight percent of students were happy with digital marking, while a smaller proportion (62%) liked the current live marking method. Students evidently recognised the benefits, implicit in their remarks: “many teachers could mark my performance”, “my English pronunciation is properly assessed” and the assessment could be “accurate”, “fair”, “reliable”, and “objective” (Q13 - Student responses).

The results indicated that students considered DMOVA more effective than the current method to support and enhance the learning of spoken English. Eighty two percent claimed that it motivated them to learn English speaking, while 62% thought the current testing method already offered this benefit. They articulated it thus: “DMOVA could help me watch and re-watch my performance to identify my weaknesses in speaking, then I try to improve my skills”, “help me review my performance to see how I speak in the test”, “see my mistakes and fix them”, “make me feel motivated because my performance can be reviewed and I can receive teachers’ feedback on my speaking”,

and “provide me accurate assessment, which motivates me to enhance my English communication skills” (Q13 - Student responses).

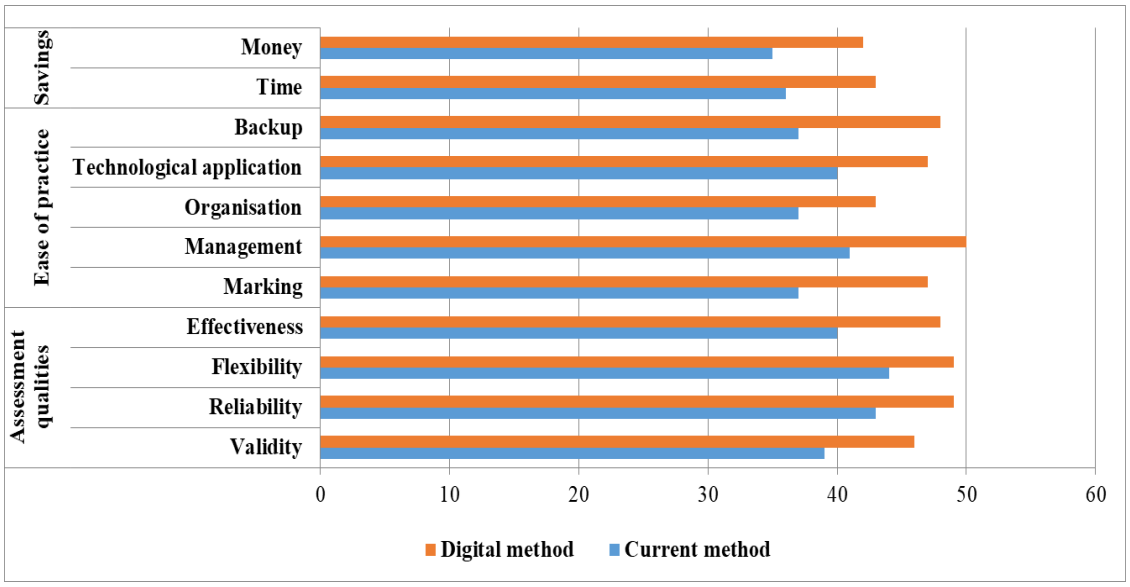


Figure 5.16 Student Perceptions of DMOVA and Current Assessment Method.

Overall, DMOVA was perceived as an effective tool for assessing speaking performance. Eighty percent of students agreed, while 67% thought the current method was effective. Other factors relating to the digital testing method, such as reliability and validity, saving money, technology use, setup time, test organisation, ease of use, flexibility, and compatibility with available resources all achieved higher-level responses than the current method.

Although the survey results identified little student dissatisfaction with the two testing methods, there were some noteworthy differences in their perceptions. Students were most unhappy about issues of cost associated with the digital testing method and expressed concerns about the expense of investing in technology and equipment. They also suggested that the digital testing method be introduced in their English course so that they could get used to the procedure and enhance their performance (Q17 - Student responses).

Student dissatisfaction with the absence of backups and the low pedagogical impact of the current testing method was evident in the data. They were also concerned about other aspects of the current testing method, such as reliability of the test results and the general effectiveness of the method.

Summary

In summary, the data analysis of the student survey highlighted the following findings:

- The majority of them had experience with computer-assisted EFL tests,
- Of all the English skills, speaking and writing were the least tested with computer assistance,
- Digital representation of speaking performances was perceived to be beneficial for assessment and learning purposes,
- Students were positive about the reliability and feasibility of DMOVA,
- Students were enthusiastic about the capacity of the digital testing method to bring about more equitable and comprehensive assessment,
- Student satisfaction rated higher for DMOVA than the current testing method.

The findings of the student survey analysis in Phase 2 aligned with the findings of the teacher survey in Phase 2 in the following respects:

- Teachers and students were persuaded by the effectiveness of DMOVA for English speaking assessment,
- Both cohorts acknowledged the benefits of DMOVA for enhancing reliability, flexibility, accuracy and comprehensiveness in speaking assessments,
- Both groups recognised the potential for DMOVA to enhance motivation and positively impact on teaching and learning,
- Overall, they were happier with benefits that DMOVA provided than the current method.

As with the teacher findings, the findings of the Phase 2 student survey also confirmed those of Phase 1. Both indicated that students were familiar and had experience with computer-assisted tests. At the time the research was conducted, computer-assisted tests for English speaking skills at FPT University were virtually non-existent. However, both surveys showed that the students responded positively to the advantages of computer-assisted tests for assessing English speaking skills. Further findings are presented in the analysis of the observation data.

Observation Data

Observations were conducted over a total of six hours, equivalent to three testing sessions. Each student was observed twice, once in the group task and again in the individual task. Observational data were noted as codes on the observation sheets.

Teacher Observations

Changes in Teacher Practice

None of the teachers observed (Teacher 1, 2, 3, and 4) had any problems with the presence of the camera in the test room. Teacher 1 confidently helped operate the OVA App on the iPad. In testing session one, she appeared to be a little nervous when asked to assist with recording videos on the iPad because it was her first experience; however, in testing session two, she was visibly more confident and less stressed. In testing session three, she took complete control of the App and the iPad and smoothly captured the performances.

Table 5.7

Teacher and Student Observation Schedule

Test session	**Teachers**	**English Level**	**Number**	**Test session**	**Teachers**
1	Teachers 1,4	Intermediate	23	46	03.04.2018
2	Teachers 1,3	Pre-Intermediate	17	34	04.04.2018
3	Teachers 1,2	High-Intermediate	20	40	06.04.2018

Teacher 1 and Teacher 4 invigilated testing session one. They appeared quite stressed in the first 30 minutes but were more relaxed by the end of the session. Teacher 1 seemed more stressed than Teacher 4, likely due to her having more responsibility for both sound and visual quality, since Teacher 1 was mainly operating the OVA App on the iPad. Teacher 4 did her usual job of invigilation and seemed more relaxed and unfazed by the camera.

Teacher 1 and Teacher 3 invigilated testing session two. Teacher 1 appeared relaxed, but Teacher 3 seemed a little stressed at the start. The test setting was formal and students were more serious than usual because they were being videoed; this may have affected Teacher 3's composure. She was observed grappling with the test procedure and operating the OVA App on the iPad but was more relaxed after a discussion with Teacher 1.

Teacher 1 and Teacher 2 invigilated testing session three. Both teachers appeared confident and relaxed. They seemed unaffected by the presence of the camera or the researcher who was sitting in the far corner of the classroom. The test was invigilated smoothly and in relaxed fashion. Although Teacher 2 had not previously been exposed to the new testing method, she did not seem stressed or flustered by the camera or video recordings.

Over the three testing sessions it became evident that teachers were changing their behaviours in relation to operating the camera and delivering the digital test. Teacher 1 was visibly less stressed and more confident after she became used to the camera in the second and third testing sessions. Teachers 2, 3 and 4 were more relaxed after the first group of students finished their performances. The researcher witnesses a positive change in teachers' behaviours – they were optimistic about the digital testing method.

Teacher Adaptation to DMOVA

Teachers were observed setting up the digital equipment in the test room. In testing session one, it took teachers and the researcher 14 minutes to complete, including a short trial recording to check sound and visual quality and adjusting the furniture. In testing session two, it took around five-and-a-half minutes to complete. Teacher 1 was responsible for setting up the digital equipment and Teacher 3 arranged the desks and chairs for the test. In testing session three, the classroom setup took two teachers just under six minutes to complete, with similar teacher roles as the second session. They were able to manage setup of the room and the digital equipment without assistance from IT or other staff.

Operating the camera was mainly undertaken by Teacher 1. She initially displayed some nervousness with the technology but overcame her anxiety by the second and the third testing sessions and encountered no difficulties operating the equipment.

For the group assessment tasks, teachers divided students into groups of four from a randomly ordered name list. After the first group had completed their test, the second group entered the test room and the teachers accommodated them effortlessly. They guided students to sit in the correct position at the desk in readiness for the test, and gave each student a card, with a number ranging from 1 to 4, to assist identification. The researcher did not observe any difficulties with the way the two teachers organised the group tasks in any of the testing sessions.

The researcher also noted the teacher instructions before the test. Each teacher took turns giving short, clear instructions related to the test questions and the time available for preparation and discussion. Teacher 1 reminded students that their performance would be videoed for research purposes. After the test, teachers briefly moderated the student results. After the last student left the test room, the two teachers compared their marking sheets, made calculations and quickly came to an agreement about the results. The average time for moderating the testing sessions was approximately three minutes, during which there was little discussion among the teachers.

Observations of the test organisation uncovered some noteworthy findings. The time for setting up the test room reduced significantly from 14 minutes to approximately five minutes in the second and third sessions. Teacher 1, who was mainly responsible for operating the camera, quickly learnt how to use the technology and subsequently experienced no difficulties. There were no issues related to organising the group tasks. The teacher instructions were clear and brief despite vast differences between the digital and current testing methods. The time for moderation was short, at an average of only three minutes per class of 20 students.

Technical Issues

No problems were observed in relation to Wi-Fi connection, software errors or video breakdowns during the three test sessions. In test session one, after a trial recording of the first group, Teacher 1 and Teacher 4 discovered that the sound recording wasn't clear enough and solved the problem by placing the camera closer to the students to improve the sound quality. They measured the distance from the camera to the student and shared this information with the other invigilators.

During all three test sessions, Teacher 1 checked the camera to ensure that it fully captured the individuals and groups of students. No issues related to the iPads or the App were observed during the three testing sessions.

Summary

Analysis of the teacher observations highlighted the following:

- There were positive changes in teacher practice and delivery of digital assessment,
- The teachers organised themselves quickly for tests using DMOVA,
- No technical issues were observed.

The data showed that the teachers were confident delivering the test using digital technology. Although they were observed being a little confused and stressed in the first few minutes, they quickly gained confidence and took control of the technologies. Despite being the first tests using DMOVA in a real testing setting, no technical issues arose and no support was needed from IT or other staff.

Student Observations

Student observations were obtained in two ways. They were observed in the test room during testing time and in the videos after conclusion of the tests. Observational data were coded on the student observation sheets and analysed using theme coding.

Student Attitudes

Sixty students were observed in three classes and each class was allocated one test session. Every student was observed twice, in an individual task assessment and a group task. Table 5.7 illustrates the student numbers and observations in each class.

The observational data in Figure 5.17 indicates that students who were confident in front of the camera and had positive attitudes toward DMOVA outnumbered those who were shy and nervous. Those with high-intermediate English appeared to be the most confident, with 62% of them unstressed by the video camera. Sixty one percent of intermediate students and fifty six percent of pre-intermediate students were confident. These students were completely engaged in their assessment tasks and seemed unaware of the presence of the camera.

The results suggest that students with higher levels of English were more confident in front of the camera, while those with lower levels of English were less confident. Pre-intermediate students were also more nervous and distracted by their surroundings than high-intermediate and intermediate students.

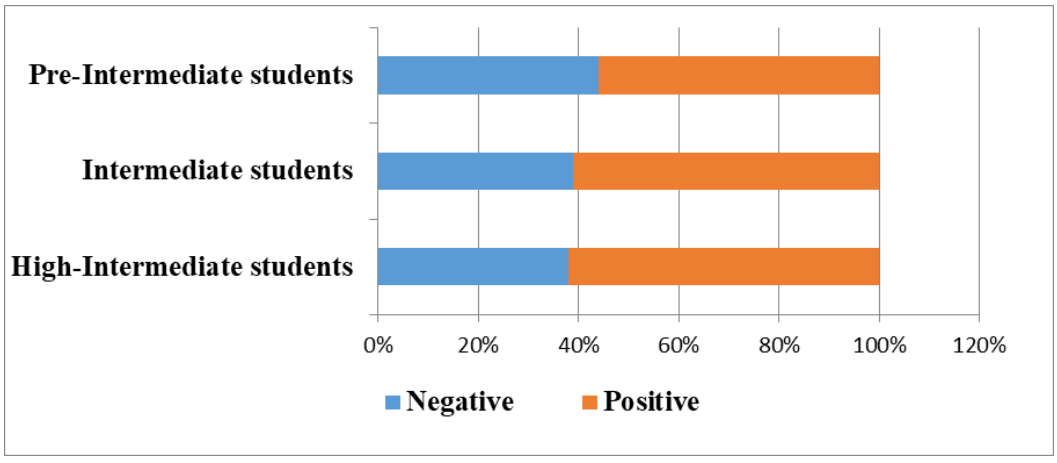


Figure 5.17 Student Attitudes Toward DMOVA.

Confident students were easy to identify in the observations. They spoke loudly and clearly without looking at the camera, were engaged in their assessment tasks, delivered their talks naturally, and spoke fluently and competently without long pauses. They had

an abundance of ideas and used expansive vocabulary in their presentations. The other students were shy and nervous and kept looking at the camera during their presentations, clearly aware of its presence in the room. They appeared uncomfortable as they adjusted their posture. One student clapped his hands with relief when the group finished their assessment task. This group of students were hesitant in their delivery and frequently looked down or sat uncomfortably while they were talking.

The graph in figure 5.18 shows the observational data of student behaviours and attitudes in each assessment task. As can be seen, the number of confident students at high-intermediate and intermediate levels was higher than those who were shy and nervous. Supported by the findings from the teacher interviews, high-intermediate students displayed more confidence in the group tasks than individual tasks. Teacher 2, who invigilated the high-intermediate class, claimed these students felt like they were acting together in a film while their performance was being videoed and were motivated to perform better as a group than as individuals.

Observations of the intermediate students showed a different scenario. These students seemed more confident in their individual assessment tasks. The group task was their first experience with the new testing technique and they were nervous and shy about being videoed. A comparatively larger number of students were concerned about the presence of the camera.

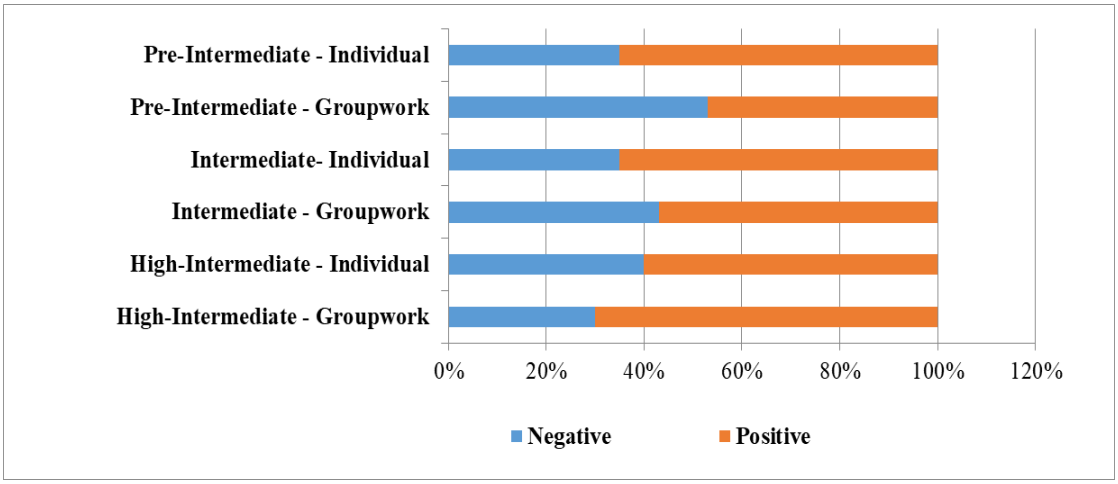


Figure 5.18 Student Attitudes Observed in Each Assessment Task.

However, their behaviours changed in the second assessment task. Students were singled out to complete their individual tasks and were seen to be more confident and engaged, taking no notice of the camera. They were more familiar with the camera and the new testing regime and their attitudes appeared more positive.

Pre-intermediate students were shy and nervous. In the group task, the number of students who were stressed was higher than those who were confident. Some students recovered from their initial nervousness and became more confident, but others remained anxious throughout. The pre-intermediate students were new to both the digital testing method and group assessment tasks, and the teachers explained that their relatively poor EFL speaking skills heightened their stress and anxiety. In their individual tasks, the pre-intermediate students displayed more confidence. They were familiar with individual assessments, having been exposed to them at beginner level, and were seen to be more familiar with the camera in the room. Eleven students were confident and comfortable delivering their talks, did not pay attention to the camera and engaged more in their tasks. Although many pauses and stops were observed in their individual presentations, the teachers attributed this to their low competence levels.

In summary, the observational data showed there were more confident students in front of the camera than nervous and shy ones. Confidence was linked to English proficiency, with more competent students displaying more confidence than the less competent students. Students were more confident in the individual assessments than the group assessments, while those with higher levels of English appeared more motivated in the group tasks.

Student Cooperation and Engagement

In the observations, all the students followed their teachers' instructions and rules in the test room. There was no evidence of cheating or disrespect in any of the three test sessions. All students participated seriously and made an effort to complete their assessment tasks. No students appeared to have difficulty getting involved in the discussion and cooperating with other group members. One or two group members were dominant over the others, for example, a high-intermediate student (S0012) in group 3 was observed supporting the other members in his group and giving them opportunities to discuss and express their ideas.

As noted, high-intermediate students engaged more fully in assessment tasks than intermediates and pre-intermediates. Eighteen high-intermediate students (18/20) were observed making and effort and concentrating on the test questions in the individual tasks. Sixteen (16/20) were absorbed in discussion and undistracted by the camera. Fifteen out of 23 intermediate students were undistracted by the presence of the camera in their group task. Fourteen students diligently completed their individual tasks regardless of the camera, seemingly oblivious to its presence in the room.

The pre-intermediate group exhibited the lowest level of engagement in assessment tasks. They continuously looked at the camera and were obviously distracted by its presence, appearing shy and nervous. Four students engaged in the group task. The others were somewhat disinterested, speaking and contributing little. Seven students conscientiously addressed the individual task. Most of the pre-intermediate students had poor English speaking skills, so their individual talks were punctuated by long pauses. According to Teacher 3, also the class teacher, this was not related to stress, but rather to their weak speaking skills and lack of English vocabulary and expressions.

All students cooperated with teachers and their peers in the group tasks to successfully complete the test. Their engagement in the assessment tasks was largely dependent on their English competence. The more competent they were, the more they engaged with the test. The high-intermediate students were more engaged and less distracted by their surroundings than the pre-intermediate students.

Time for Assessment Tasks

Although the time allowance for each assessment task was pre-set in the OVA App, students' start and finish times varied greatly. There were 16 video recordings of group tasks and 60 videos of individual tasks (see Table 5.8). Most students completed in less than the six minutes assigned for the group task and less than the three minutes assigned for the individual task.

Table 5.8

Number of Video Recordings

Class	**Number of students**	**Number of recordings**	
		Group	**Individual**
Pre-Intermediate - Top Notch 2	17	5	17
Intermediate - Top Notch 3	23	6	23
High-Intermediate - Summit 1	20	5	20

The average time duration of high-intermediate group performances was between four and six minutes, longer than intermediate and pre-intermediate students. Although some pre-intermediate groups went over five minutes, there were several long pauses during their presentations. The time duration for individual tasks varied greatly. Most high-intermediate students talked for more than two minutes, while most of the intermediates and pre-intermediates talked for less than two minutes. A few pre-intermediate students took three minutes to finish their individual presentations, but typically, with long pauses throughout. The time duration for individual tasks varied most among the

intermediate students, with the majority completing the task in one to one-and-a-half minutes. Unlike the pre-intermediate students, the intermediate students tended to conclude their presentations when they ran out of ideas.

In summary, the actual time taken to complete assessment tasks varied widely. Students with higher levels of English spoke for a longer time than those with lower levels of competence. No students complained about the time duration for the assessment tasks but recommended the OVA App contain a timer to help them better manage their time allowances (Student survey, 2018).

Summary

In general, the observations attested that the presence of the camera in the test room did not affect the usual performance of the students and supports the findings of the student survey in Phase 2 as follows:

- Surveyed students were familiar with computer-assisted tests at university
- The majority of surveyed students had previous experience with computer-assisted EFL tests.

Although some students were a little nervous to start with, they soon gained confidence. Most were unfazed by the presence of the camera. There were no apparent differences in the attitudes of students who took the tests in the current way and those who followed the digital method. They were observed focusing on the assessment tasks at hand and appeared determined to perform better, and some students reported being motivated by the digital testing method. All cooperated with their teachers and peers by engaging in the group tasks and following the test rules. There were no technical issues observed during the three testing sessions.

The data highlighted that the students' English competence contributed greatly to their confidence; the more competent they were, the more confidently they performed, regardless of the testing method.

Teacher Interview Data

Seven teachers, coded T1 to T7, participated in the semi-structured interviews. T1, T2, T3, and T4 also participated as test invigilators and markers of student digital presentations. Interviews were conducted after all teachers had finished their marking.

Interviews were conducted in a friendly environment, either in the classroom before class time or the staff room at lunch time. Teachers were also invited to talk to the

researcher during the break, with the purpose of exploring their perspectives and experiences with DMOVA in greater detail. The environment was expected to reassure teachers so that they felt free to share their thoughts and express their opinions, with the intention of eliciting the richest possible information from the interviews. Table 5.9 shows the dates and times of the teacher interviews.

Table 5.9

Teacher Interview Dates and Times

Teachers	Codes	Interview dates and times	Interview duration (minutes)
Teacher 1	T1	9:22 am, 16 April 2018	37
Teacher 2	T2	9:37 am, 19 April 2018	33
Teacher 3	T3	9:39 am, 17 April 2018	24
Teacher 4	T4	9:28 am, 19 April 2018	22
Teacher 5	T5	9:50 am, 18 April 2018	15
Teacher 6	T6	9:08 am, 19 April 2018	18
Teacher 7	T7	1:14 pm, 18 April 2018	20

After the interview data were coded using NVivo 12.1.0 the relationships between codes were identified. Significant aspects, including feasibility dimensions; digital marking and testing versus the current method; teacher acceptance and recommendations highlighted the emerging themes. The feasibility dimension covered fairness, reliability, validity, manageability, pedagogical impacts and technology.

Teacher Perceptions of Feasibility Dimensions

Based on the feasibility framework (see Figure 2.7) in Chapter 2, aspects of the functionality, manageability, pedagogy and technology of the digital method were further explored through teachers' perceptions.

Fairness

The majority of teachers agreed that DMOVA enhanced the fairness of assessment in relation to equal test times, objective and accurate marking, fair feedback, and consistency in their judgements. The findings on fairness are summarised in Table 5.10.

Table 5.10

Enhanced Fairness in Assessment

Aspects	**Strategies to enhance fairness**	**Possible enhancement**
Equal test times	Advance time setting for each assessment task	No differences in time of performance between competent and incompetent students. More similarity with writing and reading tests in terms of time
Reduction of subjectivity in marking	Invisible markers for video marking	Less distraction and interferences. Enhanced objective scoring.
Accuracy in marking	Multiple marking Review	More accuracy in marking.
Fairness of feedback	Recording feedback in the system then delivering to individual students	More accurate feedback. Fostering self-reflection based on feedback.
Consistency in teacher judgements	Replaying videos when marking for consistency in judgement. Delaying marking when feeling tired for quality of judgement.	More reliable and accurate scoring. Enhanced fairness in assessment.

In the interviews, three teachers (T3, T5, and T7) talked about fairness as an advantage of the digital method in assessing student speaking skills. Teacher 3 claimed the digital method put speaking tests on a more equal footing with reading and writing because students had more time to finish their tests, compared to the current method where students were frequently interrupted by teachers. As for tests of other language skills, the new method gave students all the time assigned and all had the same amount of time for their presentations, thereby enhancing the fairness of the process.

Teacher 3 added that the new testing method helped reduce subjectivity in marking. She reported that students often complained about disparities in marking by different teachers in the current method; some had even noticed differences in results awarded by easy-going versus serious teachers. The current testing method allowed one or two teachers to mark student performances only once in real time, with a higher risk of discrepancies. Students believed their assessments were distorted by teachers' personal judgments and their results depended on individual standards. Teacher 3 was hopeful that the digital method, which allowed multiple marking and review, would solve students' concerns in these regards.

Teacher 5 claimed that the digital testing method engendered fairer assessment because teachers were more focused on their marking. When she marked digitally, she did not

have to spend time organising the test room, grouping students or completing paperwork. Nor was she distracted by student attitudes or appearances. In addition, all students were considered equal in front of the camera and the recorded performances were carefully assessed and reassessed upon request. Teacher 5 said that she found marking the digital presentations "impersonal" (T5, Interview), which she clarified to mean that her emotions did not affect her assessment.

In the interview, Teacher 5 talked about students receiving instant feedback and suggestions in the current testing method. However, this could be viewed as a disadvantage by students who received less feedback than others. In contrast, the digital method provided students with their test results and the teachers' comments printed on paper or via email directly to the individual and not in front of the class. This was viewed as a positive approach because it prevented shame and embarrassment for the weaker students.

Teacher 7 also raised the issue of fairness with the digital testing method. He restated the benefit of being able to move back and forth over the videos as he was marking, and although this took more time, it contributed to consistency and fairness of his assessments. The risk with the current method was that the quality of marking was initially high but could deteriorate. As alluded to by Teacher 7, marking tended to become more subjective when teachers were tired. With the digital method, teachers could stop and start marking at their convenience, and in this way, DMOVA sowed the seeds for higher levels of fairness.

In summary, the teachers agreed that DMOVA offered higher levels of fairness in relation to time and marking of student performances. All students had the same amount of time for their presentations. The marking disparities between different teachers were narrowed and teacher assessments were more consistent and objective. The teachers also believed that students were treated equally when performing in front of the camera and received equal feedback and comments.

Reliability

Many teachers mentioned reliability as a strength of the new testing method. Reliability was perceived to be enhanced by accurate and consistent marking. The findings are summarised in Table 5.11.

Table 5.11

Enhanced Reliability in Assessment

Aspects	Strategies to enhance reliability	Possible enhancement
Accuracy in marking	Multiple marking Reviewing Reflecting Comparing and contrasting Onscreen digital marking key	More reliability in marking.
Consistency in marking	Focusing on marking Avoiding fatigue and distraction	Less variability in results among multiple markers.

Teacher 3 was confident that the new testing method was reliable. Although every teacher had different standards of judgement, DMOVA provided multiple opportunities for marking and review after comparing and contrasting, to narrow the gaps in results. In her view, the new testing method helped teachers focus more on their marking without being distracted by their surroundings or student behaviours and appearances, and therefore enhanced consistency and reliability. Teacher 7 also agreed that DMOVA improved marking quality by mitigating fatigue.

Teacher 4 agreed that the new testing method was more reliable than the current one, mainly due to the digital marking key embedded in the OVA App always on display next to the video, and clear criteria that simplified grading to the mere click of a button. According to Teacher 4, this function allowed her to mark more accurately by being able to refer to the marking key while observing the video. The App gave her a running total and total marks for student achievement, which she could adjust for accuracy and fairness. She complained about having to add up the points for each section to arrive at a total in the current method, and the difficulties of only knowing the total mark once the marking was done. DMOVA continually displayed the total mark and gave her more time for comparison.

Teacher 4 recommended the marking key contain more grades for each criterion to provide additional choices and more precise descriptions of student competence.

In summary, teachers were buoyant about the capacity of the digital testing method to enhance the consistency of their assessments.

Validity

Teacher 1 related the story of a high-intermediate student to whom she awarded high marks in the old testing method. When she re-marked the test using the digital method, she discovered that although the student spoke English fluently and dominated the

group, his ideas and answers were not always directly related to the questions. She immediately recognised her tendency to give the student higher marks, claiming that the digital method forced her to focus on what was supposed to be marked.

Teacher 2 found that strictly following the criteria in the DMOVA marking key improved the validity and accuracy of her assessments. “Teachers cannot be lazy and they have to mark every small criterion in the marking key objectively” (T2, Interview). She argued that teachers marked student performances more diligently with the digital method and measured what they were supposed to measure.

Teacher 3 reiterated the praise of others for the accuracy of the digital method. After her experience with digital marking, she realised that she needed to bring more objectivity to her marking in the current system. She became aware that DMOVA had reduced her subjectivity, and in turn, enhanced the accuracy of her assessments.

Teacher 4 was persuaded by the validity of the new testing method because she could measure what she was supposed to measure. She liked the clarity of the criteria in the marking key and found that she marked the videos in a more detailed manner. She added that she used analytical marking in the current testing method but a holistic approach in her final judgement, far less detailed than the analytical marking in the digital method. Most teachers concurred that digital testing enhanced the validity of assessments by encouraging them to mark according to the marking criteria and being more careful and objective. They believed that DMOVA offered more accurate outcomes because it focused their efforts on measuring what was supposed to be measured. The findings on validity are summarised in Table 5.12.

Table 5.12

Validity of Assessment

Aspects	**Strategies to ensure validity**	**Possible enhancement**
Criterion-oriented validity	Onscreen digital marking key Marking key adapted from the one currently used at the target university and IELTS public version.	Objectivity and reliability
Content validity	Reviewing and self-reflection on marking Digital marking key ensures adherence to what should be measured.	Accuracy: Mark what was supposed to be marked.
Construct validity	Clarified marking key criteria Quality videos used with the OVA App offering full functions of reviewing and peer-marking. Analytical marking	Accuracy and consistency

Manageability

The teachers were asked for their opinion on how the digital testing method supported results management and distribution, and its impact on test organisation and setup. The findings are summarised in Table 5.13.

Table 5.13

Enhanced Manageability

Aspects	**Strategies to facilitate management**	**Possible enhancement**
Test result management	Digitising and recording assessment evidence. Digitising the process of submitting results, sending performance to teachers for marking and reviewing. Onscreen marking. Saving results in the system digitally.	Enhancing professionalism. Enhancing reliability. Enhancing fairness.
Test result distribution	Digitally extracting results and feedback onto paper. Digitally sending results to related individuals. Digitally retrieving results from the system.	Saving time. Enhancing transparency.
Management of test organisation and setup	Organising the test room easily. Facilitating time management by using assessment tasks with pre-set time. Recording the contexts of performance. Not requiring technical support. Free from technical issues.	Saving time. Enhancing fairness. Reducing cheating and nepotism.

Teacher 1 made the comment that managing digital tests eliminated significant administrative labour in the current manual system and saved time by transferring the results to paper. As far as test-room management was concerned, she found the technology made it easier for teachers to manage and organise tests.

Teacher 3 had similar views about test-room management. She reported that digital assessment helped her to manage the time effectively. Having a pre-set time for each presentation helped students plan their performances to fit the timeframe, whereas the current testing method relied upon teachers using their watches or phones. Moreover, some students were allowed to keep talking after their time was up and teachers did not always interrupt them. Some teachers also prompted students with guiding questions, taking up their speaking time and advantaging some more than others.

Teacher 3 used the online timer on her smartphone to time student presentations in the current method. However, she encountered difficulties setting and managing the time; manual time setting did not work effectively when students talked enthusiastically and she was unable to stop them. In her opinion, students were more motivated to plan their performances and use their time allotment productively in the digital testing method. Teachers could also manage tests with a high degree of professionalism and accuracy. Teacher 3 had no difficulties with the technology and believed the digital method was feasible, given their IT literacy and the university's existing facilities. She found the camera easy to operate because it was not hand held for recording but set down in an unobtrusive position. The absence of any evidence of student performances in the current testing method was described by Teacher 3 as unsupportive of the assessment process. For her, recording the tests represented a step towards the same testing protocols as the other English language skills. She added that digital testing also helped manage other aspects of the test, such as minimising cheating and nepotism.

Teacher 2 agreed that the new testing method enhanced the management of speaking tests and effectively mitigated against cheating. Teacher 7 was pleased that he could plan time to mark and therefore manage his time better. Overall, teachers expressed satisfaction with the management support provided by digital assessment and frequently mentioned the advantages of managing time, technology and test rooms.

Pedagogy

The majority of teachers expected digital assessment to have both positive and negative pedagogical impacts. In the interviews, they put forward suggestions for enhancing pedagogical impact and the quality of assessments. According to most, DMOVA boosted student learning and encouraged them to practise speaking at home. It also motivated teachers to reflect on their marking. Teacher 1 observed the digital testing method increased student motivation to work on their speaking, both in class and at home. Once DMOVA was applied in practice, she encouraged students to record their own speaking performances, review them, and reflect on their pronunciation and expressions.

Teacher 2 was surprised by her students' reactions in front of the camera. Some performed much better than usual, possibly because they knew other teachers would review their videos. A few students told her that they felt motivated to perform better – she believed that the video recordings raised their awareness of how they looked and spoke on camera. In the group task, when the whole group of students were in front of

the camera, they said they felt like actors in a movie. Teacher 2 observed some of her usually quiet students being more active and confident in front of the camera. She claimed these students were very shy in face-to-face situations but spoke English very fluently when their performance was being recorded. In her opinion, the students who were partial to social networking seemed to be more confident and knew how to position themselves in front of the camera; therefore, they gave a better performance than their usual practice in English class. By contrast, some other students did not perform well because they were self-conscious and concerned about how they appeared on video. This could have undermined their confidence and negatively affected their performance. For this reason, Teacher 2 proposed that digital representation should not contain videos of the students, because some were clearly uncomfortable in front of the camera. She argued that teachers might be distracted by the students' body language but admitted that the visual aspect was essential to ensure the veracity and authenticity of the tests.

Teacher 7 also expressed concerns about the potential for visual distractions to affect marking. However, he acknowledged that the visual element was necessary to assess student delivery of their presentations, adding that it depended on the purpose of the test whether teachers should focus on listening to the audio or watching the video.

Teacher 3 was confident about the ability of the new testing method to enhance fairness and reliability in speaking tests, recognising that students would be motivated to improve their speaking. They could no longer learn topics by heart and rely on luck or prepare answers in advance to anticipated questions. Teacher 3 hoped that DMOVA would encourage the teaching of speaking skills in the same way as other language skills and encourage students to take it more seriously. She observed students trying harder when their performances were videoed and assumed they gave it their best shot because they were aware that the videos would be viewed and rechecked. Most of the students in her class said they did not feel uncomfortable or under pressure in front of the camera. Teacher 3 reported that many of her students said they liked the new testing method. She emphasised the benefit of DMOVA in allowing students to review their own performances so they could learn from their mistakes. After using the digital method for marking speaking skills, she reflected on her own practice and realised that she needed to mark more analytically by using a marking key. She also recognised a need to be more objective and avoid being distracted by external factors and personal relationships.

Teacher 1 discovered that she needed to change the way she marked student interviews. The digital marking exercise made her realise that she should focus more on her marking. She admitted that she always maintained eye contact with students when they performed, often nodding in agreement with what they were saying to reassure them. However, she recognised that continuous eye contact may have affected her concentration on what the students were saying rather than marking their competency.

In comparing the marking of interviews with that of videos, Teacher 1 acknowledged that the digital method helped her focus on listening to what students were saying, hence she was able to more accurately assess their speaking skills. By listening, she was undistracted by other factors, such as student attitudes, eye contact, and her own reactions. She said:

> I didn't recognise how much I was affected by students' attitudes and eye contact until I marked the videos of their performance. After I marked a student's video, I recognised how easily I gave him such a high mark for such a bad performance when I marked his performance face-to-face. (Teacher 1, Interview, 2018)

In summary, the majority of teachers (4) viewed the positive pedagogical impacts as an important benefit of the new testing method. The findings on pedagogy are summarised in Table 5.14. The overarching impact of the digital testing method on learning was the motivation it gave students to perform better, because the new regime, with video recording and multiple test review, elevated speaking tests to the same level of importance and fairness as other English skills tests. As a result, students were enthused to learn and practise speaking English to improve their communicative competence. Teacher practice was also positively changed, as they were obliged to teach spoken English more seriously. They had opportunities to remark student performances and reflect on their own marking. However, some teachers were concerned about the small number of students who were not confident taking tests in front of a camera.

Table 5.14

Pedagogical Dimension

Aspects	Strategies to foster EFL teaching and learning	Possible enhancement
Washback on spoken English learning.	Inspiring students' "acting" abilities in front of the camera. Encouraging students to video record their performance for review and self-reflection.	Positive impact on students' learning toward real speaking competence. Positive impact on student speaking test performances.
Washback on spoken English teaching.	Motivating teachers to teach EFL speaking. Facilitating teachers' self-reflection on their marking.	More attention to be paid to teaching of spoken English. Enhancing accuracy, reliability and fairness in marking.

Technology

Most of the teachers (4) cited the advantages and disadvantages of technology in the digital testing method and made suggestions for improving the quality of the sound recordings and reducing setup time.

Teacher 1 found the technology uncomplicated, saying that it was simple and easy for teachers to use an iPad to video the students, and the process did not require any technical support or advanced IT literacy. She participated in the study as both a test invigilator and marker and reported hardly any difference between watching the audio-visuals on video and watching students in face-to-face interviews. She said "The quality of the audio and visuals are good. The recordings are the same as the reality" (Teacher 1, Interview, 2018). Teacher 1 highlighted the important advantage of the technology's independence of Wi-Fi for averting technical problems. Although the university had good Wi-Fi transmission, teachers still experienced interruptions on occasions.

She acknowledged that teachers became distracted and tired after long periods of concentration and may sometimes miss important aspects of student presentations. In this regard, the video recordings were a useful tool for later review, thereby enhancing the accuracy of assessments. Teacher 1 had concerns about forgetting to press the START record button on the OVA App, because she had forgotten to record a pre-intermediate performance that required the student to retake the test. She suggested that teachers be carefully trained before using the equipment.

Teacher 3 declared: "This testing method was demonstrated in my class. I saw that this method was practised smoothly without any technical problems. … The technology was easy to use and could be applied on a large scale" (Teacher 3, Interview, 2018). She

found the setup and management of the test uncomplicated and did not require advanced knowledge of Information Technology. The position of the camera in the test room (see Figure 5.19) was found to be appropriate, with the camera mounted on an adjustable stand so that it didn't need handholding. Teacher 3 did not observe any problems for students caused by the presence of the camera or other technological devices.

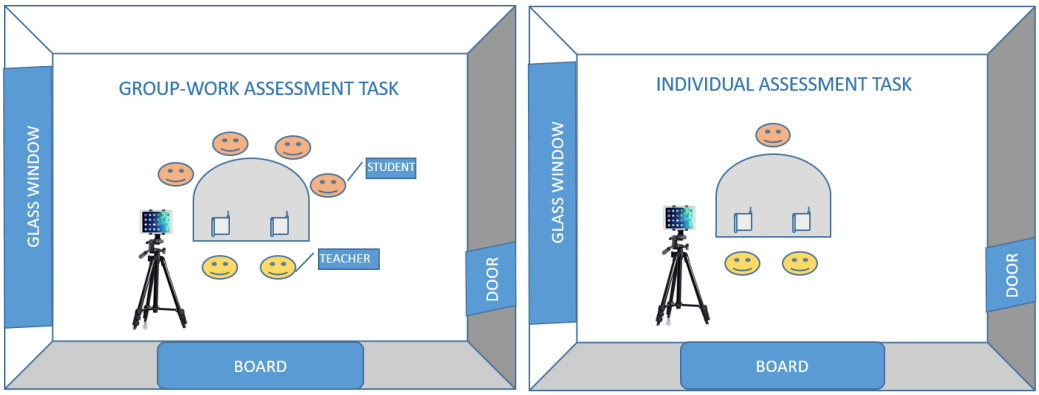


Figure 5.19 Test Room Layout.

Teacher 4 reiterated the simplicity of the new technology, claiming that the digital testing could be undertaken by anyone who invigilated speaking tests, not just English teachers: "When I do the invigilation of an English-speaking test, I merely take notes and give final assessment" (Teacher 4, Interview, 2018). She hoped that this technology for capturing student performances digitally would alleviate the need for only English teachers to invigilate English speaking tests.

Teacher 4 was satisfied with the sound quality that was improved with headphones and experienced no problems with either the audio or visual quality of the recordings. She liked the fast-forward and rewind functions of the OVA App which assisted her marking and saved time. Moreover, the technology gave her flexibility in terms of marking times and locations and she didn't have to "tie" herself to one place for lengthy periods of time. Teacher 4 was concerned about the risk of overusing the fast-forward function in the face of tight deadlines, because important aspects of student performances could be missed and potentially compromise the assessment.

Teacher 5 commented on the affordability of the technology. She proposed a better quality iPad with a reliable sound recorder for obtaining superior quality sound recordings. Coupled with being unable to clearly see the students' faces in the videos, making it difficult for her to lip-read when she didn't understand what they were saying, the sound quality left room for improvement. She suggested adjusting the camera angle

to help solve this problem. This teacher's biggest concern was that students would feel uncomfortable about speaking to a machine instead of a person and may therefore not perform as naturally as in face-to-face interviews.

Overall, most teachers were satisfied with the ease and simplicity of the technology involved in digital assessment. The findings are summarised in Table 5.15. They agreed that the technology was simple and effective for assessments and offered a variety of functions to assist their marking and manage student performances. They mentioned some disadvantages and suggested solutions, including teacher training and upgrading the technology to help solve relevant issues.

Table 5.15

Technological Dimension

Aspects	**Technical advantages**	**Technical disadvantages**
Ease of use	Easy to use. Do not require special technical support or advanced IT literacy in users.	Provide training for teachers to avoid missing records.
Usefulness	Capture high quality videos. Work efficiently for long periods of time, unlike humans. Adapt to available technologies.	Upgrade technologies for better video quality.
Innovation	Wi-Fi independence. Onscreen marking. Mobile marking.	Overuse of fast-forward function when under time pressure.

Digital Marking Versus Current Marking

Figure 5.20 illustrates the differences between the digital and current marking processes. The current method involved using paper and pencils, teachers were required to be present for the tests and mark student performances at the same time, followed by manual data entry for management and distribution purposes. In contrast, the digital method allowed teachers to access the online repository to download student performances at home and mark them using the OVA App. The results and teacher comments were automatically saved and allowed a single performance to be marked by different teachers at different times.

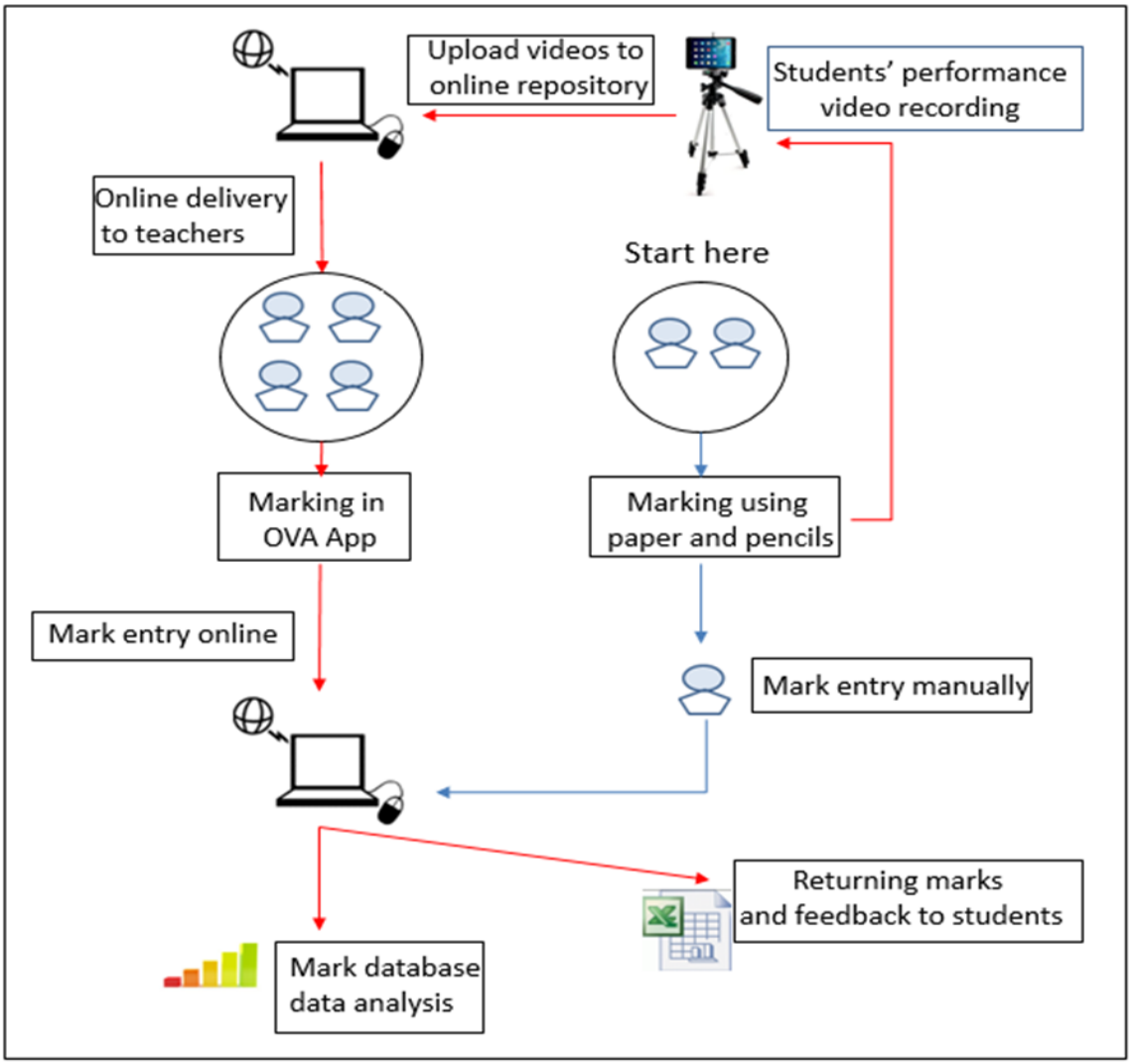


Figure 5.20 The Marking Workflow.

Digital Marking Process

After hands-on experience with digital marking, teachers were interviewed to elicit their opinions about DMOVA and their recommendations for further enhancements. They all agreed that there were both advantages and limitations to digital marking.

Advantages

Most teachers (6/7) claimed that digital marking helped them concentrate more on how students were speaking and what they were saying. They were more focused and therefore less distracted by external factors. They liked the fast-forward and rewind features for careful and accurate marking. Teacher 2 said: "I can manage students' performance by fast forwarding parts where students have long pauses. I also can rewind parts that I cannot hear clearly. I like these functions of the digital

representation." (Teacher 2, Interview, 2018). Teachers were confident that the digital method generated more reliable results, and thus enhanced the quality of assessments.

Teachers shared the view that they could mark the digital performances more analytically. According to Teacher 2, digital marking meant that teachers had to follow the marking key criteria to assess student skills. She said: "Scientifically, I find that this assessment method increases the accuracy of English-speaking assessment. Teachers cannot be lazy. They need to follow all the criteria in the marking key displayed just in front of them on the screen" (Teacher 2, Interview, 2018). Teacher 3 also reported that the marking key in the OVA App was effective in aiding analytical marking. Compared to the current marking method, Teacher 4 was partial to the clearly defined, detailed criteria of DMOVA for facilitating analytical marking.

Unlike Teacher 3, Teacher 5 used a combination of analytical and holistic marking. She found that she focused more on the content of the presentations using the digital method and was able to recognise students' weaknesses and identify areas for improvement. Teacher 5 claimed that marking with DMOVA was more "impersonal" than direct interviews but admitted being frequently distracted by students' mannerisms in direct interviews.

Teacher 6 reinforced the potential of the digital assessment method to mark more accurately, citing the ability of teachers to listen to student performances multiple times and compare students within groups to ensure fair and accurate assessments. Teacher 7 liked the flexibility of being able to plan his time for marking. In his view, digital marking ensured assessment quality from the first performance to the last, because teachers could avoid fatigue and distractions. He agreed that the new testing method allowed for more accurate assessment due to the multiple review feature and analytical marking assisted by a marking key.

Limitations

Most teachers (5/7) reported that digital marking took longer than the current method, particularly the group assessment tasks, because they had to replay the video four times to mark each member of the group. They also commented on their inability to give students instant feedback with DMOVA: "Using this testing method, I cannot give students my instant feedback. I only can write my comments in the OVA App" (Teacher 4, Interview, 2018).

Teacher 2 was distracted by students' body language when she marked digitally. In her view, the students made too many unnecessary gestures which she found distracting, a limitation of both methods. She suggested that teachers focus more on listening to what students were saying rather than watching them perform. Teacher 2 also referred to the group assessments taking longer to mark than the face-to-face interviews because she had to replay the videos several times to mark all the members of the group.

Teacher 5 suggested that students read their questions out loud at the beginning of each video. In this way, teachers would know what the questions were without referring to the question list. She was satisfied with the video quality but recommended upgrading the voice recording equipment to improve the sound quality.

Overall, teachers were dissatisfied with the time taken to mark assessments digitally, particularly the group tasks, and the lack of instant feedback. It was noted that the digital method did not completely eliminate distractions.

Current Marking Process

Three teachers agreed that the current testing method allowed them to interact with students in real time and provide students with instant feedback and suggestions (Teacher 4, Interview, 2018). The current method was effective for students with lower levels of English competence, because teachers could prompt them with guiding questions and ask them to clarify what they meant. Teachers also appeared to lipread when they couldn't hear what students were saying (Teacher 5, Interview, 2018).

Six teachers complained about the subjectivity of the current marking process. They claimed they were affected by student attitudes and inclined to award higher marks when they spoke with confidence (Teachers 1, 3, 4, 5 and 6, Interview, 2018). Furthermore, teachers had different standards of judgement, so the same performance could yield different results (Teacher 3, and 4, Interview, 2018) from different teachers (Teacher 3 and 4, Interview, 2018). Teacher 3 testified that some students believed their speaking test results depended on luck rather than competence.

Teachers mainly used holistic marking in the direct interviews (Teacher 1, 3, and 4, Interview, 2018). "Teachers tend to give estimated results when marking in the current way" (Teacher 1, Interview, 2018). Teacher 3 said she did not use detailed criteria and gave students high marks if they performed particularly well, both in their individual and group tasks. She did not believe that the current marking process with paper and

pencils encouraged teachers to mark analytically, because the marking key, printed on paper, was not always clear and teachers had to memorise all the criteria.

ENGLISH SPEAKING TEST – MARKING PAPER

Level: Date: _____ Room: ___

No	Student ID	NAME	Group-work task: 12 marks					Individual task: 8 marks						
			Fluency	Pronun-ciation	Accu-racy	Lang & Expression	Sub-Total	Fluency	Pronun-ciation	Lang & Expression	Content	Sub-Total	Total	Student's Signature
			3	2	3	4	12	2	2	2	2	8	20	
1														
2														
3														
4														
5														
6														

Figure 5.21 Marking Sheet for Current Assessment Process.

Teacher 3 reported that time limitations and an onerous workload led many teachers to skip allocating marks for each criterion and merely award an overall mark for each task before adding the totals for an overall final result (see Figure 5.21). "Obviously, giving the total marks is inaccurate and subjective" (Teacher 3, Interview, 2018). She found the digital process encouraged her to mark more analytically because the marking key was clearly displayed on the computer screen alongside the videos (see Figures 5.22 and 5.23). Teachers simply clicked on the relevant criteria and the computer calculated the results.

Figure 5.22 Marking Interface of OVA App – Individual Task.

Figure 5.23 Marking Interface of OVA App – Group Task.

Five teachers reported being easily distracted when marking interviews. Teacher 1 said: "Teachers are affected by different factors" and: "Although students' English-speaking competence was not good enough, if they showed positive attitudes and a can-do spirit, I would give them higher marks". Eye contact encouraged some students to perform, while others were uncomfortable when teachers kept looking at them while they were performing.

Teacher 5 testified that she was influenced by her personal impressions of students. In direct interviews she was frequently swayed by their efforts to deliver their presentations and was inclined to be more generous in her judgement. She added that the ability of teachers to do thorough and accurate assessments was compromised when they were tired.

Three teachers noted that marking interviews was stressful and tiring (Teacher, 2, 3, and 7, Interview, 2018). In a two-hour English-speaking invigilation with 20 students, Teacher 2 managed to concentrate on marking the first 10 but felt "overloaded" by the rest. As her fatigue increased, her concentration decreased. She explained that a huge amount of information needed to be analysed and assessed in a relatively short period of time, and her assessments after the first 10 students were not as rigorous and accurate because she was too tired to make appropriate judgements.

Teacher 3 also found the digital method helped ease marking. Marking interviews required teachers to concentrate for long periods of time and she often felt stressed and tired. She discovered that she tended to assess more subjectively when she was tired after long stretches of concentrating and didn't hear as clearly. Teacher 3 suggested that two or more teachers mark student interviews to avoid missing any aspects of their performance, but without the recordings of student performances, she was concerned about nepotism and cheating.

Teacher 7 agreed that the quality of marking interviews was likely to be higher at the start of the session than at the end. He said, "I could hardly concentrate at the end of the testing session. I was too tired". He restated the risk of increased subjectivity when fatigued.

Teacher 5 was concerned about perceptions of unfairness in the interviews, when teachers prompted some students with guiding questions to help them along, but not others. Since the number of guiding questions was randomly determined by individual

teachers and varied for each student, this practice could raise issues of inequality amongst students.

The teachers cited both advantages and disadvantages of the current marking method. On the positive side it encouraged teacher and student interaction, and teachers were able to provide students with instant feedback. On the negative side, the following issues were raised:

- Assessments were more likely to be subjective,
- Teachers' judgements were affected by both internal and external factors, for instance, students' mannerisms and teachers' personal feelings and impressions,
- Teachers experienced fatigue and stress when they had to assess a large class of students and concentrate for long periods of time,
- There was a risk that teachers might miss parts of student performances due to distraction and fatigue,
- The current method did not encourage teachers to mark analytically,
- Without recordings of student performances there was no opportunity for review,
- Teachers' prompting some students could be perceived as inequitable.

Table 5.16 summarises the key findings from the teacher interviews regarding the advantages and limitations of the digital and current marking methods.

Table 5.16

Pros and Cons of Digital and Current Marking Methods

Current marking method		**Digital marking method**	
Advantages (+)	**Limitations (-)**	**Advantages (+)**	**Limitations (-)**
Teachers could: - Provide instant feedback and suggestions. This method supported teacher-student interaction. It was effective for students with low levels of English.	Teachers could: - Mark subjectively without detailed criteria. - Easily be distracted while marking. This method generated inconsistencies in teachers' judgement.	Teachers could: - Concentrate on what was supposed to be marked and reduce distraction and fatigue. - Mark more analytically. - Mark accurately. - Mark flexibly in	Teachers: - Could not provide instant feedback. - Took more time, especially marking group tasks. - Were still distracted by students' body language. This method did not include test questions in the videos.

Digital Versus Current Assessment Process

Digital Assessment

Advantages

The majority of teachers viewed the recordings of student presentations and the backup they provided as an advantage of the digital method. Teacher 1 restated the benefits of being able to review student performances to check the results or revise their marking. She claimed that, in the interview testing method, she sometimes awarded students higher marks than they deserved. With digital marking, she could check and review any aspects of student performances if she was unsure of her initial judgment.

Teacher 3 attributed students' diligent approach to their speaking tests to being recorded. They were aware that their performances would be reviewed and remarked by other teachers and were motivated to perform better. She also mentioned that the recordings would help prevent cheating and nepotism, and therefore enhance fairness. Teacher 5 was pleased with the flexibility offered by the digital marking method in terms of time and location for marking and liked that teachers could mark from home using the videos instead of attending and observing interviews.

Five teachers expressed satisfaction with the ease of using the new testing method. Teacher 1 said: "This testing method is quite easy and convenient to apply" (Interview, 2018), adding that setup of the test room with all the required technology was simple and quick and the technology was easy to operate. Teacher 3 found the digital testing method easy to use and apply on a large scale and claimed that it reduced her workload with regard to time setting and calculating total marks. She said: "This method might make my invigilation easier and less stressful" (Teacher 3, Interview 2018).

Most teachers (5/7) recognised the benefits of the digital method in supporting invigilation and backup, exempting them from close observation, real-time marking and having to provide immediate feedback. They believed that the digital testing method did not need to be invigilated by EFL teachers and could be undertaken by any staff, potentially resolving the shortage of EFL teachers. Teacher 2 agreed that this method of marking saved time. She enjoyed having total control of the digital representations and the ability to fast forward, rewind, pause and stop as required. She also agreed that these types of assessments did not require EFL teachers to invigilate, as long as a staff member was available to operate the camera.

Most teachers (5) expressed the view that digital assessment offered more reliable and accurate test results (Teacher 1, 2, 3, 4, and 5) by reducing subjectivity as "a long step in enhancing accuracy" (Teacher 1, Interview, 2018). Teacher 1 stated it reduced distractions associated with interviews.

Teacher 2 defined fairness as providing every student with accurate assessments. Since digital representation allowed for multiple marking and review of student performances, the test results were more likely to be accurate. Five teachers concurred that equal test times for all students was a positive aspect of the digital assessment process. Teacher 2 was pleased that it reinstated equal performance times for all students.

Teachers recognised the positive impacts of digital assessment on learning and testing. Three (Teacher 1, 2, and 3) found their students were motivated to perform better and made more effort when they knew their performance was being recorded. Some of Teacher 2's students surprised her with their speaking competence and confidence in front of the camera, telling her that they paid more attention to their body language and tried to use appropriate gestures in the videos. For this teacher, the digital method facilitated formative testing to check student learning and provide them with ongoing feedback. In addition, it supported test administration and was therefore also suitable for summative tests.

The teachers highlighted six advantages of the digital assessment method as follows:

- Back-up for review and revision
- Allows multiple marking and review
- Enhances fairness, reliability and accuracy of assessment
- Flexible in terms of assessment time, location and staff
- Easy to use
- Generates positive impact on EFL speaking learning and assessment.

Teachers acknowledged that the technology could be applied on a large scale because it was easy to use, did not require high levels of IT competence, and was compatible with current university facilities.

Limitations

Some teachers observed students being nervous in front of the camera: "My students were not familiar with video recording in the speaking test because they hadn't attended a test like this before" (Teacher 5, Interview, 2018). "Some students were not confident

with their own appearance in the test with video recording" and "What would I look like in the videos?" (Teacher 2, Interview, 2018).

Teacher 2 detected a hidden fear among students in the digital test. Although it employed the same marking key as the current test and was invigilated by teachers who were familiar to them, students appeared anxious about other teachers who may mark their videos:

> One of my students told me that performing in front of the camera, she did not know who was marking her performance, and how that teacher felt about her speaking and she could not observe the teachers' facial expressions to adjust her speaking. She suddenly felt worried and was afraid that her performance would be assessed more rigorously. (Teacher 2, Interview, 2018)

The lack of teacher-student interaction in individual assessment tasks was raised as one of the limitations of the digital method. In individual interviews, teachers sometimes acted as interlocutors, prompting students with guiding questions to assist them. However, it was found to be more suitable for group assessment tasks, characterised by student-student interaction.

Nervousness in front of the camera and the fear of being judged by unknown teachers were identified as limitations of the new method for students. It was also viewed as obstructing teacher-student interaction in individual assessment tasks.

The advantages of the digital process, as perceived by teachers, far outnumbered the limitations. The benefit of backing up performances gave teachers more flexibility and enhanced reliability by allowing review and multiple marking. DMOVA did not require EFL teachers to invigilate speaking tests. It was viewed as a source of motivation for students to learn speaking and improve the quality of their performances in tests. However, teachers observed some of their students feeling nervous and self-conscious about their appearance in front of the camera and suggested that the new method may be more suitable for group tasks which involved no teacher-student interaction.

Most teachers expressed acceptance of the digital assessment method and concurred that it had the potential to enhance the quality of speaking assessment. They saw it as an effective method that significantly changed the way teachers assessed speaking skills and motivated students to learn and improve their assessment tasks. Teacher 2 said: "I totally support the digitisation of EFL speaking assessment" and: "Hopefully, this testing method will be applied successfully. If it is applied in practice now, it will surely

make significant changes to the way we are assessing EFL speaking" (Teacher 3, Interview, 2018).

Advantages of the Current Assessment Method

Three teachers talked about the benefits of the current testing method. Teacher 1 commented that in the interviews, teachers and students made eye contact and teachers could observe students' speaking and confidence levels. She believed that a positive approach deserved recognition even when students hadn't mastered their speaking skills, stating: "Even though the student's speaking is not very good, he speaks with an attitude of making an effort, trying for improvement, and cooperation, I will give him higher marks" (Teacher 1, Interview 2018).

Teacher 2 found the current testing method more authentic and said it facilitated teacher and student interaction. In the face-to-face EFL speaking tests, she explained that some female students took their cues from teachers' facial expressions and adjusted their delivery accordingly to obtain the best results for their performance.

Teacher 5 cited a student's comment about obtaining support from teachers in the interviews as a benefit of the current method. She defined "support" in the speaking tests as guiding questions and teachers' instructions for students to repeat words or sentences that were not clearly heard or understood. She believed this kind of support helped and encouraged students with their presentations.

In summary, teacher and student interaction was considered the main benefit of the current testing method. Teachers could observe students' efforts in real time and assist them with prompts and guiding questions to encourage them and for which they were duly rewarded.

Limitations of the Current Assessment Process

Most teachers reported being frequently distracted by students' appearances and attitudes, test room facilities, and their own state of mind (Teacher 1) when they invigilated speaking tests. Teacher 2 said that a two-hour testing session exhausted her, so she became easily distracted. Teacher 3 sometimes invigilated three speaking test sessions with around 20 to 25 students in one day, each lasting two hours. She was tired and thirsty but unable to leave because she was the only invigilator present. Teacher 3 had difficulty managing the time for each student's talk – three minutes for individual tasks and six minutes for group tasks – and although she set the time on her phone, students continued talking when their time was up.

Teacher 2 commented on the shortage of EFL teachers, which meant there was sometimes only one invigilator in the test room. In such cases, no moderation occurred and the invigilator's decision was final. Nor were there any recordings of student performances for later review, so these assessments tended to be subjective and the results dependent on one teacher's judgement. Teacher 2 also recognised inequalities associated with the guiding questions. Teachers who asked fewer questions at the end of the test sessions because of time pressure did not give those students the same opportunity to develop and enhance their speaking. It was apparent from their feedback that teachers mainly focused on listening in the latter part of the testing sessions and reduced their questions to students.

Teacher 5 acknowledged the inconsistencies in teacher assessments, mainly due to exhaustion towards the end of the testing sessions. According to her, these inconsistencies resulted in unfair and unreliable assessments. Table 5.17 presents the key findings from the teacher interviews regarding the advantages and limitations of both digital and current assessment processes.

Table 5.17

Comparison of Digital and Current Assessment Processes – Teacher Perspectives

Current assessment process		**Digital assessment process**	
Advantages (+)	**Limitations (-)**	**Advantages (+)**	**Limitations (-)**
Teacher and student interaction. Helped teachers observe students' speaking manner. Allowed teachers to give students instant feedback.	Easily distracted teachers. Long working hours tired teachers. No moderation if one invigilator present. No recordings of students' performance for backup and review. Did not mitigate against cheating and nepotism.	Facilitated recordings and backup Supported review, remarking and reflection. Motivated students to perform better. Mitigated against cheating and nepotism. Was easy to practice. Did not require EFL teachers to invigilate tests. Offered reliability, accuracy, fairness and flexibility to assessment process. Reduced subjectivity.	Students may feel nervous in front of the camera. May have a hidden fear of invisible markers. Lacked student and teacher interaction.

Teachers praised the current testing method for its authentic interaction, eye contact, visible facial expressions, and support with guiding questions to clarify pronunciation. On the other hand, they criticised the current testing method for being subjective and personal, inherent distractions, and inconsistent assessment

Teacher Recommendations and Suggestions

Marking Key

The marking key used in this research was digitised and functioned as a spreadsheet. Although it was adapted directly from the one the university was using, teachers made some recommendations for improvements. Teacher 2 acknowledged that the digital marking key had advantages over the paper one but maintained that methods both had their limitations. She recommended that the grades be further calibrated for each criterion because she sometimes had difficulty awarding a mark when she felt students deserved a middle mark. Teacher 1 suggested that each criterion be accompanied by a brief description for quick and easy reference.

Marking Interface of the OVA App

Teacher 1 proposed changing the marking interface for group tasks to facilitate marking and reduce marking time. Teacher 2 suggested that the names of each student be visible in the group task videos so that teachers could mark all group members in one sitting.

Information Security

Teacher 3 drew attention to information security when the recordings of student tests were uploaded to the internet for marking.

Audio or video or Both?

Teachers 2 and 7 questioned whether the students should be captured on audio or video or both. They explained that they focused only on listening to the videos and therefore found the visual aspect unnecessary. Teacher 2 did however concede that the visual element played an important role in preventing cheating and ensuring that only authorised students participated in the test. Teacher 5 reported that the visual aspect of the videos was useful for marking the way students delivered their speech. He resolved that the decision to use audio or video or both should depend on the purpose of the assessment and teachers should have the freedom to decide.

Summary

Analysis of the teacher data showed that DMOVA was believed to enhance the fairness, reliability and validity of English speaking assessments. The teachers acknowledged that the digital method facilitated management of tests and test results and had a positive pedagogical impact on both student learning and teacher practice. They expressed the view that the technology required for digital assessment was easy to use and required no technical support. The presence of the technology in the test rooms did not appear to cause any undue issues for teachers or students. The findings from the teacher interview data are summarised in Table 5.18.

Table 5.18

Feasibility of The Digital Assessment Method

Attributes	**Current assessment method**	**Digital assessment method**
Fairness	Influenced by students' attitudes and appearance. Feedback provided inequitably.	Reduced distraction and subjectivity. Enhanced fairness. Consistent judgement.
Reliability	Marking was done once. There were no recordings of student presentations.	Multiple marking and review generated consistent, precise and reliable results. Analytical marking followed the marking key and enhanced accuracy and consistency.
Validity	Teacher and student interaction was more authentic. Overall judgement was applied. Marking was not done analytically.	Enhanced validity of EFL speaking assessment. Teachers concentrated on marking what was supposed to be marked. Enhanced attention to detail in marking.
Manageability	Marking, distributing and retrieval of test results were all done manually. Did not support the management and recording of test evidence.	Assisted management and distribution of results. Improved time management and enhanced professionalism of assessment. Prevented cheating and nepotism.
Pedagogy	Students memorised a list of topics in preparation for the tests. Distractions decreased teachers' focus on marking. Did not allow for teachers' to review or reflect on their marking.	Encouraged students to practise their English speaking. Motivated students to perform better. Allowed students to review and recheck their performance and learn from their mistakes. Helped teachers reflect on and improve their marking
Technology	Did not require technology.	The iPad was easy to use. The camera captured the videos effectively for marking. The technology is Wi-Fi independent. Improved the quality of assessments in terms of providing backup, enabling review and enhancing accuracy. Did not require IT support or high levels of IT literacy. Did not cause any serious problems for teachers or students.

The findings from the teacher interviews confirmed the findings from the other data sources, viz., the teacher survey in Phase 2, teacher observations and student observations. The findings on the benefits of the digital testing method from the teacher survey in Phase 2 are restated as follows:

- The quality of assessments was enhanced by improved reliability, validity, fairness, and flexibility,
- Backup of student performances was valuable for multiple marking, review, reflection and learning,
- Motivated improved teaching practices and student learning,
- Facilitated managing assessments and was compatible with existing technologies,
- Encouraged analytical marking,
- Generated positive impacts on English testing, teaching and learning.

The findings from the teacher and student observation data attested to the following:

- Teachers adapted quickly to the digital testing method,
- No technical problems arose during the test sessions,
- There were more confident students in front of the camera than shy and nervous ones.

Analysis of the teacher interview data showed the advantages of the digital assessment process far outnumbered the limitations. Benefits included enhanced accuracy, reliability, fairness and flexibility in assessments, as well as effective test delivery, results distribution and backup. Despite the perceived limitations of some in relation to the lack of teacher and student interaction and instant feedback, the teachers expected the digital method would nevertheless enhance the quality of EFL spoken assessments and positively drive improvements in testing, learning and teaching of spoken English.

Test Results Database

Assessment Tasks and Scores

As previously described, each student completed two assessment tasks – both were video recorded. They were assessed by means of live and digital marking methods. Live marking was undertaken while teachers were invigilating the speaking tests, while digital marking was carried out using videos of student performances uploaded to an

online repository. Teachers were able to mark online or download the videos to their personal computers and mark offline.

Two EFL teachers invigilated and marked during the test performances, so each student received two marks for each assessment task. After all the videos were uploaded to the online repository, four teachers, including the two who did live marking, were invited to mark digitally. Accordingly, each student received four marks awarded by four different teachers. The allocation of teachers can be seen in Table 5.19.

Table 5.19

Allocation of Teachers to Marking

EFL level	**Live Marking**	**Digital Marking**
High-Intermediate	T1 + T2	T1 + T2 + T3 + T4
Intermediate	T1 + T4	T1 + T2 + T3 + T4
Pre-Intermediate	T1 + T3	T1 + T2 + T3 + T4

Three classes participated in the tests, comprised of 20 high-intermediate, 23 intermediate, and 17 pre-intermediate students, for a total of 60 altogether. High-intermediate students were learning Summit 1, intermediates were learning Top Notch 3, and pre-intermediates were learning Top Notch 2. Appendix S shows the correlations between Summit 1, Top Notch 3, and Top Notch 2 content and International Standards and Tests, including the Common European Framework (CEF), International English Language Testing System (IELTS), and Test of English as a Foreign Language (TOEFL).

Teacher Allocation for Marking

Four teachers participated in both live and digital testing of student performances; Teacher 1 (T1), Teacher 2 (T2), Teacher 3 (T3), and Teacher 4 (T4). Table 5.19 shows the role played by each teacher in the marking processes. Teacher 1 was the benchmark teacher, whose assessment was adopted as the standard judgement, as she had over 10 years' experience teaching EFL at tertiary level and had invigilated hundreds of EFL speaking tests during her career.

After invigilating and marking the student interviews, teachers were provided with recordings of the same student performances on iPads, also available online. Each teacher was assigned a unique user name and password to access and mark the digital recordings. Both the digital and live marking results were securely stored in the online

repository, administered by the administrator and developer of the App, Dr Alistair Campbell, at Edith Cowan University in Western Australia. Prior to the digital marking sessions, teachers were provided with a marker guideline (see Appendix T) showing them the steps for marking with the OVA App and the functions for exporting the results to Excel.

Marking Key

The marking key in this study was adapted from the one currently in use at FPT University, Vietnam, and the public version of the IELTS Speaking Band Descriptor (see Appendix U). It was divided into two parts: Part 1 included criteria for group task assessments, and Part 2, for individual task assessments. The total mark was 20 (100%). Group assessments accounted for 60% of the total result or 12/20, and individual assessments contributed forty percent or 8/20. Each criterion was allocated a different score depending on the weighting for each English level and assessment task and all were described in detail together with their equivalent scores.

At the time this study was conducted, one marking key was used for all three English levels: pre-intermediate, intermediate, and high-intermediate. However, the higher the English level was assessed, the higher requirements were. Detailed explanations were added to each criterion in the marking key to enable its specific use. At the start of each semester, EFL teachers attended a training session provided by the English department to update them on any changes in assessment, teaching methods and policies. The four teachers who marked the student performances were all experienced EFL teachers who had invigilated over 200 hours of EFL speaking tests between them at FPT University.

The teachers who marked live were provided with hard copies of the marking key and marking sheets (see Appendix M). The marking sheets looked very similar to the ones they currently used. Teachers had to write down scores for each criterion, obtain students' signatures confirming they had sat the test, sign to verify they invigilated and marked the test, and record any unexpected issues that arose. Based on university policy, they could decide on the penalty percentage for students who were caught cheating and could enter the reduced score into the database before distributing the results. Teachers were instructed to mark the same way they usually did when invigilating EFL speaking tests.

The marking key was incorporated in the OVA App to assist marking. Rather than using marking sheets, the digital marking key was placed alongside the video in each student

performance. The scores were displayed under each criterion; teachers simply clicked on the relevant criterion and entered a score. The OVA App added the scores automatically and displayed the grand total. The Marking Guidelines for Teachers (see Appendix T) was distributed to teachers in advance.

Descriptive Statistics and Correlation Analysis

Descriptive statistics and correlation analysis were used to explore relationships between the live and digital marking methods. Correlation analysis measured the degree of agreement between the teacher results for the current and digital marking methods and described the strength of the relationship between the two methods.

Correlations between the live and digital markings were measured, as well as between individual and group marking. The results of the analysis for each English level were compared in order to identify the English level and type of assessment task most effectively evaluated by the digital method.

The students were assigned to one of three English competency levels; the test results for each level were held in separate databases. Descriptive statistics and correlation analysis were applied to each database to identify relationship between live and digital marking and between individual and group marking.

High-Intermediate English Level

Relationship Between Live and Digital Marking

The analysis showed similar live marking scores for teachers T1 and T2, ranging from 7 to 17 and 7 to 17.5 respectively. There was a slight difference in their digital marking scores, from 8 to 15 and 8 to 17 respectively. While T1 did not award the higher top mark in the live marking, she was inclined to award slightly higher marks than T2, with an overall average of 12.85 (SD = 2.92) compared to T2 at 11.65 (SD = 2.95). By contrast, T1 assigned slightly lower marks than T2 in the digitally marked test, with overall averages of 11.55 (SD = 2.23) and 12.65 (SD = 2.49) respectively. Table 5.20 shows the descriptive statistics for the live and digital marking test scores.

Table 5.20

Descriptive Statistics on Live and Digital Marking Results

Pairs		No of students (N)	Min	Max	M	SD	Mean difference
Live marking	T1	20	7.00	17.50	12.85	2.92	
	T2	20	7.00	17.00	11.65	2.95	1.25
Digital marking	T1	20	8.00	15.00	11.55	2.23	
	T2	20	8.00	17.00	12.65	2.49	1.10
	T3	20	9.00	17.00	12.55	2.01	
	T4	20	8.00	16.00	11.75	2.14	0.80

It is likely that the differences between the live and digital marking by T1 and T2 were partly due to the digital method providing more time for teachers to mark so that they could plan their marking to avoid fatigue, stress, and overload, as articulated by T2 in the interview. It could also be related to T1's testimony that listening to the recordings multiple times allowed her to assess student speaking skills more accurately. Contrary to the interview method where she was inclined to award higher marks for positive attitudes and behaviour, she claimed not to be affected by student attitudes and behaviour when she marked digitally.

The data analysis highlighted agreement between all the markers, with slight differences in means that were higher in the live marking. The digital test results of the four teachers were very similar, with the mean difference of 1.10, lower than the mean difference of the live marking results (1.25). The digital marking method achieved a higher level of agreement than the live marking, as confirmed by the correlation analysis results (see Table 5.21).

Table 5.21

Correlations Between Live Marking and Digital Marking Results

		Live marking		Digital marking			
		T1	T2	T1	T2	T3	T4
Live Marking	T1	1					
	T2	0.77**	1				
Digital marking	T1	0.87**	0.85**	1			
	T2	0.55*	0.76**	0.65**	1		
	T3	0.52*	0.48*	0.49*	0.41	1	
	T4	0.52*	0.53*	0.50*	0.46*	0.34	1

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

According to Pearson's correlation coefficient (r) (Dancey & Reidy, 2007), this study categorised correlation levels as: weak positive for $0.10 \leq r < 0.40$, moderate positive

for $0.40 \leq r < 0.70$, and strong positive for $0.70 \leq r < 1$. In social sciences, results are considered to be significant at the level of 0.05 or less (Field, 2013).

A Pearson correlation coefficient analysis showed that live and digital marking results of all the teachers yielded a correlation coefficient mostly ranging from medium to strong positive (see Table 5.21). In the live markings, T1 and T2 produced similar results (r = 0.77**). Their digital marking results were also correlated at r = 0.65**. Overall, the analysis of T1 and T2's live and digital marking results indicated a strong correlation, with correlation coefficients of 0.87 and 0.76 respectively.

Digital marking results were also relatively correlated, ranging from weak to high positive. T3's digital result was the outlier, possibly because this was her first hands-on experience with the digital marking method, having reported in her interview that it took her some time to get used to the digital marking key and marking more consistently based on the criteria.

Aside from the teachers' experience, time constraints may also have impacted on their accord. T4, who intended to give lower scores compared to the others, reported in the interview that time constraints put pressure on her to fast forward parts of the videos and she was concerned that she may have missed important aspects of the performances.

Individual and Group Marking

In this part of the data analysis, the submarks awarded for individual and group tasks were analysed. Descriptive statistics showed similar mean scores for the live and digital marking of these two assessment tasks. Closer examination of the individual markings indicated that the four teachers' digital marking produced very similar results, with a similar range and small mean score differences.

For the individual assessment task there was a discernible difference in T2's results. She awarded the lowest mark (1) to the individual task in the live marking; however, in the digital marking, she assigned a mark of 3, similar to the other teacher's mark. This appears to confirm T2's view that the digital marking method gave rise to equal assessment by reducing her workload and allowing her to plan her marking, as she was unable to guarantee fair and accurate judgements after long periods of live marking.

The small mean score differences among teachers for the group marking were nevertheless larger than those for the individual tasks. The mean score difference for the digitally marked group task was larger than the live marking, and opposite to that of the

individual task assessment test. Based on the standard deviation results, the group tasks yielded a wider distribution of results compared to the individual tasks. This could be attributable to the perceptions of Teacher 2 and others in the survey, that the digital marking platform was not as effective for group tasks.

The results of the individual and group tasks are shown in Table 5.22 and Table 5.23. The correlation analysis indicates a significant correlation between T1 and T2's results for the individual tasks in both the live ($r = 0.61^{**}$, $p < 0.01$) and digital marking ($r = 0.71^{**}$, $p < 0.01$). The only insignificant correlation between T1 and T2's individual tasks was between T1's live and T2's digital results. Correlations between the results of T3 and T4 were somewhat varied.

Table 5.22

Correlations Between Live and Digital Marking – Individual Task

		Live marking		**Digital marking**			
		T1	**T2**	**T1**	**T2**	**T3**	**T4**
Live Marking	T1	1					
	T2	0.61**	1				
Digital marking	T1	0.67**	0.75**	1			
	T2	0.43	0.65**	0.71**	1		
	T3	0.59**	0.60**	0.43	0.42	1	
	T4	0.26	0.58**	0.41	0.32	0.25	1

** Correlation is significant at the 0.01 level (2-tailed).

The results of the live and digitally marked group tasks also produced significant correlations, except for T3's digital marking, once again likely due to her lack of experience with the digital method.

Table 5.23

Correlations Between Live and Digital Marking – Group Task

		Live Marking		**Digital Marking**			
		T1	**T2**	**T1**	**T2**	**T3**	**T4**
Live Marking	T1	1					
	T2	0.76**	1				
Digital marking	T1	0.83**	0.80**	1			
	T2	0.60**	0.74**	0.62**	1		
	T3	0.40	0.18	0.27	0.23	1	
	T4	0.59**	0.63**	0.48*	0.74**	0.33	1

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

In summary, the results of the two groups of teachers who marked both live and

digitally were very similar. There was a strong correlation between the live and digital marking methods and between the individual and group tasks. Teachers appeared to adjust their marks when they marked digitally. For instance, T1 awarded lower marks in the digital test, explaining that re-listening to the recordings and reviewing them multiple times enhanced the accuracy of her assessment. She was unaffected by other factors that might otherwise compromise her assessment.

The data also indicated that the four teachers' digital marking of individual tasks were more highly correlated than their live marking. This was the opposite way around for the group task marking, which had a lower correlation than the live marking. The teachers were of the view that the OVA App did not support group marking as effectively because they had to replay the recordings multiple times to mark each student, which took longer than the live marking.

Intermediate English Level

Relationship Between Live and Digital Marking

T1 and T4 invigilated and live marked the intermediate testing session. As shown in Table 5.24, T1 was inclined to award higher top marks than T4 in both her live and digital marking. Although the two teachers' marking patterns in both methods were quite similar, T1 assigned higher marks than T4. The mean scores showed that both teachers gave lower average marks in their digital marking, i.e., M (T1-Live marking) = 12.47 and M (T1-Digital marking) = 10.95. The difference between the two teachers' mean scores reduced when they marked digitally. The distribution of results for each marking method by teacher was similar: SD (T1-Live marking) = 2.21 and SD (T1-Digital marking) = 2.28.

Table 5.24

Descriptive Statistics for Live and Digital Marking

Pairs		**No of students (N)**	**Min**	**Max**	**M**	**SD**	**Mean difference**
Live marking	T1	23	8.00	17.00	12.47	2.21	
	T4	23	8.00	16.00	11.26	1.88	1.22
Digital marking	T1	23	8.00	16.00	10.95	2.28	
	T4	22	7.00	14.00	10.36	1.86	0.59
	T2	23	10.00	18.00	13.52	2.15	
	T3	23	9.00	17.00	11.65	2.05	1.77

Table 5.24 shows little difference between the averages and distribution of teachers' live and digital marking. A comparison of minimum, maximum and mean scores

identified that teachers had a tendency to award lower marks in their digital marking, reflective of the findings in the teacher interviews. T1 admitted she was easily influenced by her personal impressions of students' appearance, attitudes and confidence, and tended to give higher marks for displays of positive behaviours. The digital method allowed her to reflect on her live marking and apply more accurate judgements.

The correlation analysis (see Table 5.25) showed a weak correlation between T1 and T4's live marking (r = 0.32). However, their digital marking results were significantly correlated (r = 0.67**). The results of T4's live marking strongly correlated with the other three teachers' digital marking, while there was a moderate correlation between the results of T1's live marking and the other teachers' digital marking.

Table 5.25

Correlations Between Live Marking and Digital Marking

		Live marking		**Digital marking**			
		T1	**T4**	**T1**	**T2**	**T3**	**T4**
Live Marking	T1	1					
	T4	0.32	1				
Digital marking	T1	0.54**	0.74**	1			
	T2	0.59**	0.77**	0.86**	1		
	T3	0.44*	0.74**	0.94**	0.75**	1	
	T4	0.43*	0.70**	0.66**	0.53**	0.67**	1

* Correlation is significant at the 0.05 level (2- tailed).
** Correlation is significant at the 0.01 level (2-tailed).

The highest correlation was between the results of T1 and T3's digital marking (r = 0.94**) and the lowest correlation was between the results of T1 and T4's live marking (r = 0.32). The correlation analysis verified a significant correlation between T1, T2, T3 and T4's digital results, ranging from medium to high positive.

Individual and Group Task Marking

The data showed somewhat diverse top and bottom marks for both individual and group assessments tests. The digitally marked individual results showed that teachers were inclined to raise the minimum and lower the maximum scores, which was the opposite in the digitally marked group tests, where the mean scores for live and digital marking of individual tasks were similar, but those for group tasks varied. The mean scores of all the results for both live and digital marking were similar. The small mean and standard deviation differences suggested that teachers marked fairly consistently, regardless of the method.

The results of T1 and T4's live marking of individual tasks correlated significantly at the strong positive level (r = 0.89**), as did the results of their digital marking (r = 0.79**) (see Table 5.26). Their results for individual tasks were also significantly and strongly correlated with those of T1, T2, T3 and T4's digital marking. Again, the analysis signalled a strong correlation between teachers' live and digital marking of individual tasks, ranging from moderate to strong positive.

Table 5.26

Correlations Between Live and Digital Marking – Individual Task

		Live marking		**Digital marking**			
		T1	**T4**	**T1**	**T2**	**T3**	**T4**
Live Marking	T1	1					
	T4	0.89**	1				
Digital marking	T1	0.90**	0.90**	1			
	T2	0.70**	0.81**	0.81**	1		
	T3	0.74**	0.81**	0.87**	0.67**	1	
	T4	0.76**	0.79**	0.76**	0.58**	0.66**	1

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Similarly, correlations were noted between T1 and T4's live and digital marking of the group task at r = 0.50* and r = 0.76** respectively (see Table 5.27). Digital marking was more correlated than live marking. The results of T1 and T4's live marking correlated with those of T1, T2, T3 and T4's digital marking, spanning a range between moderate and strong positive. While the results of all four teachers' digital marks yielded correlations, they were diverse, ranging from weak positive (r = 0.37) to strong positive (r = 0.93**).

Both the live and digital marking of students' individual tasks yielded higher correlations than those of the group tasks marked by the same teachers in the same way. The digital results of all four teachers for individual tasks showed significant correlations at the 0.01 level. However, the digital results of the group task varied, with a weak positive and moderate positive response. The analysis suggested that individual assessments may be more suitable for the digital marking method than group assessments.

Table 5.27

Correlations Between Live and Digital Marking – Group-work Task

		Live marking		Digital marking			
		T1	**T4**	**T1**	**T2**	**T3**	**T4**
Live Marking	T1	1					
	T4	0.50*	1				
Digital marking	T1	0.70**	0.79**	1			
	T2	0.52**	0.47*	0.48*	1		
	T3	0.70**	0.78**	0.93**	0.37	1	
	T4	0.53**	0.76**	0.65**	0.65**	0.62**	1

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

In summary, there were no significant differences between the teachers' results for live and digital marking; they remained consistent throughout the assessment of the entire group of students. However, similar to the analysis of high-intermediate students, the study identified a tendency by teachers to award lower results to the same student's digital presentation. Further examination also revealed that digital marking yielded a higher correlation than live marking.

The submarks indicated that the results of individual assessments enjoyed higher correlations than the group tasks marked by the same teachers using the same marking methods. This finding echoed the high-intermediate cohort analysis, suggesting that the digital testing may be more effective for individual assessments than group tasks.

Pre-Intermediate Level

Relationship Between Live and Digital Marking

The descriptive statistics described similar results for T1 and T3's live marking. These teachers gave the same lowest and top mark: 6.00 and 15.00 respectively (see Table 5.28), and their mean scores and standard deviations were similar. However, the digital marking showed diverse results. The two teachers gave different lowest and top marks; with the lowest marks 4.00 and 6.00 respectively and the top marks 11.00 and 14.00 respectively. Mean scores were lower than for their live marking, suggesting that these teachers tended to give lower results for digital assessments.

Distribution of the digital results for T1 and T3 were narrower (SD (T1) = 1.74 and SD (T3) = 2.20) than the live interviews (SD (T1) = 2.68 and SD (T3) = 2.35). The four teachers' digital marking results were distributed differently, ranging from an SD of

1.52 to 2.52, indicating that their digital marking was not as consistent as their live marking for this English level.

Table 5.28

Descriptive Statistics for Live and Digital Marking

Pairs		Number of students (N)	Min	Max	M	SD	Mean difference
Live marking	T1	17	6.00	15.00	11.70	2.68	
	T3	17	6.00	15.00	11.76	2.35	0.06
Digital marking	T1	17	5.00	11.00	8.17	1.74	
	T3	17	5.00	14.00	9.35	2.20	1.18
	T2	17	4.00	14.00	10.41	2.52	
	T4	17	6.00	13.00	9.25	1.52	1.16

Analysis (see Table 5.29) identified a strong correlation between T1 and T3's live marking results (r = 0.70**) at the 0.01 level. The correlation between their digital results was even higher, with a significantly strong reading (r = 0.92**) at the 0.01 level. T1 and T3's live marking was consistent with their digital marking, with significantly strong correlations r = 0.86** and r = 0.85** respectively at the 0.01 level. Teacher 3 attributed her disparate results between the two marking methods to enhanced objectivity in her digital assessments. She also credited the digital marking method with improving her accuracy.

Table 5.29

Correlations Between Live Marking and Digital Marking

		Live marking		Digital marking			
		T1	T3	T1	T2	T3	T4
Live Marking	T1	1					
	T3	0.70**	1				
Digital marking	T1	0.86**	0.75**	1			
	T2	0.73**	0.53**	0.65**	1		
	T3	0.80**	0.85**	0.92**	0.60*	1	
	T4	0.41	0.61*	0.37	0.54*	0.39	1

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

The results of T1's live marking significantly correlated with T2 and T3's digital marking at r = 0.73** and r = 0.80** respectively. T3's live results also correlated with the other teachers' digital marks; while T4's digital marks least correlated with the other teachers. This could perhaps be explained by her inclination to fast forward the student recordings, particularly during long pauses, with a heightened risk of missing important aspects of their presentations.

Individual and Group Task Marking

The analysis showed similar results for individual tasks marked live by T1 and T3. It also showed that the other teachers' digital marking was lower than their live marking. Although there was an apparent tendency among teachers to award lower marks when they marked digitally, their marking was consistent, with similar mean scores and small standard deviations.

Compared to individual tasks, the group task results were also lower in the digital assessment, and were adjusted down by teachers, generating larger gaps in mean scores. The data analysis suggested that teachers made numerous adjustments to group results when they marked digitally. The results reflected Teacher 1's comments about her tendency to award higher marks when she marked student performances live. She blamed students' appearance and other distractions, such as eye contact, their disposition, and cooperation. When she marked digitally she was unaffected by these factors and able to concentrate on what was supposed to be assessed.

Significant correlations were identified between the individual tasks marked live and digitally by the four teachers. T1 and T3's live marking of individual tasks showed a significantly strong correlation, r = 0.71** at the 0.01 level (see Table 5.30). The results of these two teachers' live marking correlated significantly with the digital results of the others, within the moderately significant to strongly significant range.

Table 5.30

Correlations Between Live and Digital Marking – Individual Task

		Live marking		**Digital marking**			
		T1	**T3**	**T1**	**T2**	**T3**	**T4**
Live Marking	T1	1					
	T3	0.71**	1				
Digital marking	T1	0.84**	0.78**	1			
	T2	0.76**	0.68**	0.79**	1		
	T3	0.80**	0.72**	0.92**	0.72**	1	
	T4	0.62**	0.61*	0.64**	0.67**	0.64**	1

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

T1 and T3's digital marks yielded a strong significant correlation, r = 0.92** at the 0.01 level; higher than the correlation between their live marks at r = 0.71**. Their digital marking of individual tasks were significantly correlated, ranging between moderately significant (r = 0.64**) and strongly significant (r = 0.92**). These two teachers' live marking of group tasks produced a moderately significant result (r = 0.59*) at the 0.05

level, and a strongly significant result (r = 89**) at the 0.01 for their digital marking. The data suggest that the adjustments made by teachers when marking digitally generated more correlated results.

Table 5.31

Correlations Between Live and Digital Marking – Group Task

		Live marking		**Digital marking**			
		T1	**T3**	**T1**	**T2**	**T3**	**T4**
Live Marking	T1	1					
	T3	0.59*	1				
Digital marking	T1	0.69**	0.66**	1			
	T2	0.52*	0.42	0.36	1		
	T3	0.69**	0.76**	0.89**	0.38	1	
	T4	0.25	0.52*	0.13	0.45	0.22	1

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

T2 and T4's digitally marked group tests correlated least with the other teachers' live and digital marking. Although the group tasks were positively correlated, most of these were either moderately significant or weakly insignificant. The group tests were less correlated than the individual tests.

In summary, the correlation coefficient of pre-intermediate student outcomes marked by different teachers using the current and digital methods unveiled four main findings. First, the correlation between the live and digital results marked by T1 and T3 was statistically significant. Second, the digital marking results of T1 and T3 were more correlated than their live marking results. Third, the correlations between the digital tests marked by the four teachers were significantly positive, with digital results lower than live test results. Fourth, the correlations between the digitally marked individual assessments were stronger than those between the group assessments marked the same way.

Summary

There was a common tendency among teachers to award lower marks for digital assessments. In spite of this, all the teachers' results for every English level assessed using the live and digital marking methods were quite similar. Analysis of the results database showed significantly positive correlations between live and digital marking at the 0.01 level (see Table 5.32).

Table 5.32

Correlations between Live and Digital Marking

	T1	T2	T3	T4
High-Intermediate	0.87**	0.76**		
Intermediate	0.54**			0.70**
Pre-Intermediate	0.86**		0.85**	

** Correlation is significant at the 0.01 level (2-tailed).

The analysis also indicated that the correlations between the digital marking results were higher than the live marking results of the same teachers (see Table 5.33). For all three English levels, the digital results identified significant positive correlations, with the highest correlation (r = 92**) in the pre-intermediate cohort. In the intermediate group of students, a significant positive correlation (r = 66**) was observed – the same teachers' live marking did not yield a significant correlation (r = 0.32).

Table 5.33

Correlations between Results Marked Live and Digitally

	T1 – T2		T1 – T3		T1 – T4	
	Live	Digital	Live	Digital	Live	Digital
High-Intermediate	0.77**	0.65**				
Intermediate					0.32	0.66**
Pre-Intermediate			0.70**	0.92**		

** Correlation is significant at the 0.01 level (2-tailed).

Correlation analysis of the submarks in the group and individual tasks marked digitally showed the individual tasks returned higher correlations among teachers than the group tasks. Descriptive statistics identified diversities in the teacher results for group tasks marked digitally. As reflected in the interviews, teachers found the OVA interface not as effective for marking group tasks because it took them longer to mark than the interviews and may suggest that DMOVA is more effective for individual than group assessments.

Conclusion

Chapter 5 presented the findings of Phase 2 of the study, aimed at answering the research questions by analysing the data collected from survey questionnaires, observations, interviews and speaking tests. The following findings emerged:

a) Teachers and students had positive perceptions of the digital assessment method.
   - Teachers and students at the university were familiar with computer-assisted EFL tests.

- Of the four English skills, speaking skills were the least assessed with computer-assisted tests.
- DMOVA was perceived to be beneficial for assessment and learning purposes.

b) Teachers had no difficulties using the digital assessment method.
   - Teachers were confident about delivering English speaking tests with digital representation.
   - No technical issues were observed in the tests using DMOVA.

c) Teachers believed that DMOVA was feasible.
   - **Fairness:** Fairness was enhanced by minimising distractions and subjectivity, thereby maintaining consistency.
   - **Reliability:** Reliability was enhanced by enabling multiple marking and review, and encouraging analytical marking by adhering to a marking key for consistent, precise and reliable results.
   - **Validity:** The validity of assessment was enhanced by inducing more detailed and careful marking.
   - **Manageability:** The workload associated with storage, distribution and management of the results was minimised by the digital process, at the same time elevating English speaking assessments to a new level of professionalism.
   - **Pedagogy:** Students were motivated to perform better, review their presentations and learn from their mistakes. Teachers could reflect on their marking and improve their assessment skills.
   - **Technology:** Implementation and operation did not involve costly investment or require IT support and high levels of IT literacy.

d) The results of the live marking correlated significantly with those for digital marking.
   - Analysis implied that teachers marked consistently, regardless of marking method.
   - Correlations between the digital marking results were higher than the live marking results of the same teachers.
   - The digital results for all three English levels returned significant positive correlations.
   - Across all three English levels, the results of the individual tasks showed higher correlations than the group tasks marked by the same teachers.

The findings of both Phase 1 and Phase 2 of the study are further explained and evaluated in Chapter 6. Relationships between the findings, the literature review and the research questions are also discussed in further detail.

CHAPTER 6
DISCUSSION OF FINDINGS

This study investigated the feasibility of implementing DMOVA for the assessment of EFL spoken language in a university context in Vietnam. As far as could be ascertained, the literature has not confirmed the use of digital representations to assess EFL spoken language on a large scale, although it has been used for assessing student performances in some subjects, such as Italian, Applied Information Technology, and Engineering in a Western Australian educational context. Despite its potential for enhancing the assessment of EFL spoken language that is in dire need of innovation and renewal, the feasibility of this testing method in a Vietnamese context has not yet been measured. It was also necessary to understand the benefits and limitations of this testing method for optimal uptake and implementation. The findings reported in the previous chapter addressed the research questions throughout and these questions are revisited below as a preface to discussing the findings.

In addressing the overarching research question: How feasible is digital representation for summative assessment of EFL speaking performance in Vietnam? this chapter is divided into three main sections; each discusses the findings in relation to the three subsidiary questions. First, the perceptions and acceptance of stakeholders are outlined, followed by the feasibility of implementing DMOVA for the assessment of spoken English. The third section discusses the benefits and limitations of implementing DMOVA in a university context in Vietnam, before the chapter concludes with a brief summary and recommendations for further studies.

Stakeholder Perceptions and Acceptance

Subquestion 1: What are teacher and student perceptions of computer-assisted EFL speaking assessment? This subquestion included three questions:

1. What language testing techniques are currently used in Vietnam?
2. What are teacher and student views of computer-assisted assessment (CAA)?
3. Do teachers and students show an attitude of willingness toward the introduction of a computer-assisted assessment trial?

In terms of language testing techniques, the survey results showed that three assessment methods were currently used at FPT university for assessing students' EFL competence:

paper-and-pencil tests, oral tests and computer-assisted language tests. An important finding was that computer-assisted English assessment was the dominant method for testing English in EFL classes. This differed from the study of Sinwongsuwat (2012), who claimed that paper-and-pencil EFL tests were still predominantly used in EFL classes to assess students' English competence in Thailand.

The current study also found that both the teacher and student participants were familiar with digital testing techniques for EFL and possessed appropriate ICT literacy levels to take on the proposed technologies for learning, teaching and testing EFL skills. These findings were verified in both phases of the study. However, they do not support previous research that indicated the use of technologies in language teaching and learning challenged students and teachers (Uzunboylu & Tuncay, 2010), and risked scaring language teachers off due to their lack of ICT training and insufficient technological knowledge and experience (Hu & McGrath, 2012; Wang, 2014).

A further finding highlighted in the first phase of the study was that the digital testing used by teachers for assessment focused mainly on listening and reading skills. It was not being used to assess English speaking, once again supporting Phase 2 of this study and previous studies in Vietnam (Canh, 2013; Hoang, 2010; Tran, 2013) and Thailand (Sinwongsuwat, 2012). In Thailand "students' communicative abilities are still assessed by means of paper-and-pencil multiple-choice tests, particularly in large-scale school and university admission exams" (Sinwongsuwat, 2012, p. 76).

In relation to computer-assisted assessment (CAA), the survey indicated that both teachers and students had positive attitudes and were confident with computer-assisted assessment. Both cohorts said they preferred this method to the current paper-and-pencil method, for several reasons. First, teachers indicated that computer-assisted English tests offered more advantages, such as immediate feedback, improved manageability, objectivity and enhanced efficiencies in terms of time and cost. Second, students believed this testing method offered them convenience in terms of time and location, immediate feedback, simplicity of use, resource efficiency, high levels of precision and fairness, and a reduction in stress levels. The positivity expressed by participants towards the use of CAA corresponds with the study by Wang (2014), who observed teachers' positive attitudes towards integrating ICT in teaching.

The current research unveiled some teachers' cynicism towards the authenticity of computer-assisted tests for EFL speaking. They were concerned about the capacity of

digital tests to offer real-life contexts as effectively as traditional testing methods, consistent with prior studies that suggested English speaking should be assessed as oral interaction in real-life contexts (Brown, 2003) and computer-assisted assessments fail to foster conversations and interactions like face-to-face interviews (Kenyon & Malabonga, 2001). Teachers were also concerned about the reliability of scoring in the computer-assisted method, given that computers were not yet capable of measuring all the richness of human speech, including nuances, turn-taking and negotiation (Moere, 2010). However, other research contradicted Moere's study and showed a high correlation between tests scored by humans and those scored by computers (Bernstein et al., 2010). The author acknowledged "one of the undoubted advantages of computer-delivered speaking tests is their high reliability due to the standardisation of test prompts and delivery, which naturally eliminates any interviewer variability" (Kenyon & Malone, 2010, p. 36). The survey results in the current study attested to teacher satisfaction with the marking reliability of face-to-face interviews, yet prior studies claimed that assessments conducted by human markers involve a great deal of subjectivity (Harmer, 2014), influenced by markers' wellbeing, tiredness, concerns and stress (Hartle, 2009).

It is possible that teachers' scepticism about the reliability and authenticity of computer-assisted EFL speaking assessment was due to their lack of practical training and experience with integrating technologies, particularly for testing EFL communicative competence. This view was expressed in both phases of the study and suggested that some teachers were reluctant to adopt the new technologies for assessing student speaking skills and hesitant to change their practice. It accords with research by Uzunboylu and Tuncay (2010), who encountered significant diversity in teachers' digital capacity, and Wang (2014), who identified a gap between teachers' expressed enjoyment of using technology and their actual use of technology in tertiary teaching.

In terms of participant support for computer-assisted assessment, both Perceived Usefulness and Perceived Ease of Use were positively identified by the technology acceptance model (F. Davis et al., 1989). Teachers and students were upbeat about using digital testing and exhibited strong Behavioural Intention to using the technology in a trial. The willingness of teachers and students to adopt the technology was consistent with a study by Zhan and Wan (2016), who found students welcomed the innovation of computer-based English listening and speaking tests. This is understandable, given the specific research context of FPT University in Vietnam,

where computer-assisted tests were frequently used for assessing EFL competence. Although there was a critical need for improving English speaking, assessments lacked integrated technologies. The surveys confirmed that both teachers and students had high levels of IT literacy. Teachers had experience with design, customisation and delivery of computer-assisted language tests and students were familiar with taking language tests on computers. Their willingness to participate in a digital EFL speaking trial signalled a desire to use modern technologies for improving communicative assessment. They expressed hopefulness in the technology to solve current assessment issues and generate positive impacts on teaching and learning.

Feasibility of Implementation

Subquestion 2: What is the feasibility of digital representation of student performances for English speaking assessment in terms of functionality, manageability, pedagogy, and technology?

Functionality

The functional dimension explored in the current study was based mainly on stakeholder perceptions of assessment validity, reliability and fairness, as well as the correlation analysis of EFL speaking test results scored digitally and live. These aspects are discussed in turn below.

Validity

After scoring, most teachers agreed that DMOVA provided a true representation of student performances. They were satisfied with the quality of the videos and confident of their capacity to enhance scoring accuracy. This finding aligns with a study by Kirkgoz (2011), who identified positive perceptions on the part of teachers towards implementing video recordings in task-based learning classrooms and recommended video as a valuable learning resource. The current study also concurs with research indicating that video recordings provide direct evidence for assessment and support reflection, peer feedback and analytical discussion (Borko et al., 2008; Rosaen et al., 2008; Santagata, 2009).

The onscreen digital marking key, adapted from the one in use at FPT University and the IELTS public version, was a key contributor to objectivity and reliability, according to the teachers. It clarified the marking criteria, thereby enhancing transparency of the assessment. The onscreen marking key also encouraged teachers to use an analytical marking method, suggesting that criterion-oriented assessments ensured validity,

consistent with the assertion of Costa and Kallick (2004), who argued that valid assessment should be based on criteria.

In addition, the digital assessment method facilitated review and self-reflection, which in turn, fostered accuracy. The digital marking key required teachers to consistently assess what was supposed to be assessed, and in so doing, enhanced content validity. Teacher reviews and reflection on their marking went a long way towards strengthening the detail, accuracy and consistency of assessments. In the current study, teachers' affirmation of validity reflected the early definition of Young and He (1998).

Across all three English levels, there was a correlation between the test results of both the digital and current marking methods. DMOVA facilitated multiple marking and review, enhancing consistency and reliability in scoring and providing feedback. The results suggested that the reliability of the scoring supported the validity of the assessment. They also confirmed that digital testing was a valid method for assessing EFL speaking. The outcomes of the English test interviews strongly correlated with the results of the digital assessments, as in other studies where the "validity argument for indirect speaking tests has been that they measure the same construct as direct speaking tests … The argument is that if scores on two tests are so highly associated that one can predict from one to the other, the test must be 'construct-equivalent'" (Fulcher, 2014, p. 172). According to Harmer's (2014) definition, the similarities between the two different methods of testing the same abilities of students demonstrated the criterion validity of DMOVA.

Factors that threatened the validity of assessments were also examined, including technical problems, confidential scoring, student confidence and teacher bias. These potential threats were foreseen and minimised during the assessments, such that there were no technical breakdowns. Teachers were provided with unique usernames and passwords to access the scoring system and maintain confidentiality. In addition, the majority of students appeared confident in front of the camera. There were therefore no visible impacts on the validity of digital assessments.

The results of the study showed that digital testing was suitable for the context of a university in Vietnam, where teachers and students possessed high levels of IT literacy and were familiar with computer-assisted EFL assessment. The university was also equipped with modern technologies that were compatible with DMOVA. For all these reasons, the digital method was appropriate for stakeholders and the context, where

higher levels of reliability and validity were needed to change the assessment of EFL spoken language for the better.

Reliability

Most teachers in the current study were convinced that DMOVA provided more reliable results than the current method, due to more accurate marking. The digital method facilitated multiple marking, peer marking, peer review, multiple review and reflection, consistent with early research that showed multiple ratings by certified teachers (Thompson, Buck, & Byrnes, 1989) increased the reliability of oral proficiency assessment. This also concurs with a more recent study of Yu (2012), who found the standardised procedures in computerised speaking tests assessed speaking more accurately than interviews.

Onscreen marking with the marking key encouraged teachers to adhere to the criteria and mark analytically. Analytical marking was credited by Barkaoui (2011) for its detailed feedback on student performances and high-level consistency. The current study suggests that DMOVA enhanced the reliability of assessments by encouraging analytical marking, as in a study by Jonsson and Svingby (2007), who proved that analytical marking using rubrics enhanced scoring reliability in performance assessments. Analytical marking can identify individual students' strengths and weaknesses (De La Paz, 2009); however, it might not be able to provide as complete a picture of student performances as a holistic measuring scheme (Moskal, 2000).

Phase 1 raised the issue of scepticism among teachers about the reliability of computer-assisted English speaking assessment, although they agreed it reduced their subjectivity. In Phase 2, teachers recognised the effectiveness of DMOVA in enhancing reliability through having more experience with DMOVA and self-reflection on their marking methods

In contrast to the teachers, Phase 1 results indicated that 99% of students found the current assessment method reliable. However, after the DMOVA trial, there was a significant change in their perceptions, with 72% satisfied with the reliability offered by digital testing. After the trial, nearly three quarters of the student cohort considered DMOVA a more reliable method of assessment than the current method.

Phase 2 results showed teachers believed DMOVA enhanced the reliability of speaking assessments in terms of accuracy and consistency in their marking. Accuracy was enhanced by the strategies employed to mark digital performances, including multiple

marking, review, reflection, comparing and contrasting, and using the digital marking key. Consistency was improved because they were able to focus on what they were supposed to mark and avoid fatigue and distractions, resulting in less variability between markers. This finding aligns with Harmer (2014), who claimed the reliability of a test is affected by the way the test is marked, and when teachers observe and assess rather than being an interlocutor, assessments are more reliable. Sundqvist, Wikström, Sandlund, and Nyroos (2018) also found that recordings of student speaking tests removed teachers from the distractions of face-to-face encounters.

Teachers' digital results attested to an increased use of analytical marking. Most teachers reported that they closely followed the onscreen marking key, resulting in them using the analytical marking method. The design of the OVA App facilitated analytical marking rather than holistic marking, as recommended for oral assessment by Harmer (2014) to enhance reliability. This suggests that analytical marking improved the reliability of the digital assessment method. Additionally, the design of the OVA App appeared to foster standardisation in teachers' marking, thereby enhancing consistency.

Reliability of digital assessment in this study was defined in terms of score equivalence between the current and digital methods, as well as the advantages of multiple marking and review offered by DMOVA. The discussion on score equivalence below looks at the types of assessment tasks that were more effectively assessed by DMOVA.

Score Equivalence

Speaking test results were collected across three levels of English competence and included two assessment tasks conducted at the end of each semester. The teachers who invigilated and marked the trial tests were experienced in these areas and used a marking key adapted from the one used by FPT University at the time of the research.

The correlation analysis showed the live and digital results for all three English levels yielded significant correlations (see Table 5.35), as did the marking of the individual and group tasks. The findings corroborated the contention of Chiedu and Omenogor (2014), who claimed that there is "a measure of reliability obtained when a language teacher creates two forms of the same test by varying the items slightly. Reliability is stated as a correlation between scores of Test 1 and Test 2" (p. 6). The score equivalence of the same test using both the digital and current methods was shown to be reliable.

Correlations in this study had parallels with the findings of Bernstein et al. (2010) and Stansfield and Kenyon (1992). In their validity study of fully automated delivery and scoring of spoken language tests, Bernstein et al. (2010) found a high correlation between scores derived from interviews and automated tests. Agreement on scores obtained from simulated interviews and live interviews was also the focus of a study by Stansfield and Kenyon (1992). The current study contributed to the literature by identifying correlations between live and digital results across different English levels in a context where English was taught and learnt as a foreign language. There was very little in the literature on correlations between assessment results generated from digital representation and the currently used assessment method for EFL. The findings confirmed significant correlations between the two assessment methods and endorsed the digital assessment method as a reliable alternative. In fact, the digital results were positively significantly correlated, while the live results yielded lower or no significant correlations (see Table 5.36), suggesting that live results were not as consistent as digital results.

In the current study, it became evident that teachers tended to award lower scores when they marked students digitally. While this may have been disappointing for EFL students, the correlations between the live and digitally marked results were significant. The findings suggest that teachers reflected on their marking practices and adjusted their assessments in digital marking. In the teacher interviews, they reported being inclined to adjust their scores for the sake of accuracy using this method, when they recognised they had overlooked something or over-evaluated a performance. The ability to re-mark and review were likely to lead to more accurate assessments of competency.

To avoid bias, all teacher participants were experienced with invigilating and marking speaking assessments. The results showed agreement between their digital scores, i.e., T1's digital marking correlated with the other three teachers. This may signal a relationship between teacher experience and marking, which, although not measured in the current study, may indicate a further means of enhancing the assessment process. L. Davis (2016), Harmer (2014) and Nyroos and Sandlund (2014) claimed that reliability is not only affected by the way tests are marked but also by the people who mark them, and teacher experience can have an effect on scoring reliability (Nyroos & Sandlund, 2014). A wider range of teachers would have to be recruited to investigate this claim further.

Multiple Marking and Review

Among the 18 teachers interviewed in Phase 2, seventeen indicated that DMOVA allowed them to mark and review student speaking performances multiple times. They commented on their heightened accuracy as a result of revisiting the videos numerous times and not missing important aspects of student performances. DMOVA also allowed multiple teachers to access the system, thereby enhancing reliability, since it encouraged peer marking, full double marking and multiple marking. This supports Harmer's (2014) claim that more than one scorer marking the same students' work can greatly enhance reliability, and aligns with Galaczi (2010), who argued that computer-delivered speaking tests enhanced reliability because they included more raters in the assessment process.

Teachers attested to improvements in the reliability of speaking assessments using DMOVA. Teacher 1 claimed in the interview that digital marking was more accurate than live marking because it was less subjective. She found that distractions in the live marking sessions diverted her attention from the content of student performances, relating how one high-intermediate student (S005) dominated the group with his strong personality and impressive manner of speaking. She awarded him 17.5/20, while another teacher scored him 12/20 (see Table 6.1), but when she re-marked the digital presentation, she realised that the student had not answered the questions satisfactorily in terms of accuracy, language, and expression. Accordingly, she adjusted her mark down to 14/20, which was the same score awarded by the other teacher for the student's digital test.

Table 6.1

High-Intermediate Student Test Results

Student	**Live T1**	**Live T2**	**Digi T1**	**Digi T2**
S005	17.5	12	14	14

The above findings show that the ability to review student performances helped teachers reflect on their marking, an aspect of the digital method that isn't possible with live marking. Teachers also articulated the drawback of having no record of tests in the current assessment method, consistent with Sundqvist et al. (2018), who showed that recording speaking tests enabled re-listening and collaborative assessment. In that study, the lack of recordings translated into having no evidence of teacher practice and raised questions about standardisation in speaking assessments (Sundqvist et al., 2018).

Fairness

The majority of EFL teachers were of the view that DMOVA enhanced the fairness of speaking assessments by fostering objective, accurate marking and feedback, and more consistent teacher judgements. This aligns with Stowell's (2004) concept of fairness, defined as consistent treatment, particularly in group tasks. Stowell (2004) argued that student performances should be fairly assessed, based on their fulfilment of assessment tasks.

In the current study, the DMOVA re-listening and review features contributed to fair assessment by enhancing the probability of equitable judgement by teachers. Additionally, DMOVA allowed teachers the freedom to mark at their convenience, potentially avoiding issues of fatigue, boredom and inconsistent marking. Their positive opinions of DMOVA's capability for multiple review and assessment mirrors Shohamy's (2000) definition of assessment fairness. The author claimed that fairness can only be assessed from several demonstrations of proficiency, such as portfolios, self and peer assessment; and a fairness assessment model is democratic and ethical about the way knowledge is assessed and the test results are used.

In this study, perceptions of fairness related to the validity and reliability of assessment. Objectivity, accurate marking, and provision of feedback were identified by participants as catalysts for positive change. In digital marking, teachers were invisible to the students. They were also free from distractions and other influences that potentially skewed their judgement, such as students' mannerisms and their own inclinations to prompt students. There was general consensus among most participants that multiple marking, listening and review opportunities contributed to the accuracy of assessment. Teachers identified the advantages of having more time to record their feedback with the digital method, ultimately enhancing both teaching and learning.

Another aspect of fairness highlighted in the current study was the equal use of test time. This meant that every assessment task was assigned a predetermined time and students were the sole users of that time in any way they chose. Equal test times were also perceived to narrow the gap between assessments of English writing, reading and speaking skills.

Manageability

As clarified in the feasibility framework (see Figure 2.7), the manageability dimension involved administering assessments, including the collection, storage and distribution of

students' work and results (Kimbell et al., 2007). In the current study, manageability was examined through the lens of participant experiences and perceptions of DMOVA in facilitating test management and results distribution. Further research on management for administrators and app developers is recommended to complete the entire picture.

In this study, most teachers agreed that DMOVA was an improvement on the conventional method for managing EFL speaking tests. The digital testing method digitised the test evidence and results before being submitted to administrators, distributed to teachers for marking and review, and saved in computer systems for subsequent retrieval. It eliminated the manual work associated with writing feedback, typing and printing results, as well as filing. DMOVA computerised the entire process by allowing the results to be exported to Excel, emailed and retrieved at the touch of a button. It was also perceived to ease the burden of organising and setting up speaking tests and required no technical assistance or support.

Onscreen marking was sparsely mentioned in the literature on computer-assisted language assessment, particularly speaking assessment; and was regarded by the teachers in this study as a highly innovative feature. They liked the analytical marking aspect, which they believed enhanced reliability and saved time. Despite being a new concept, the teachers' positive perceptions of DMOVA were evident in and from the data, echoing the findings of Coniam (2013), who reported a growing acceptance of this method among young markers in public Hong Kong examinations. The author predicted that onscreen marking would become the norm, due to strong indications of inter-rater reliability and correlations between onscreen and paper- marked scores. Given its potential contribution to consistency, onscreen marking of speaking assessments is worthy of further research. The teachers' positive perceptions of the logistical advantages for collecting, multiple marking, storing and distributing student work and results concurred with previous results reported by Kenyon and Malone (2010). Multiple marking entailed teachers being assigned unique usernames and passwords so that their results were confidential and they could evaluate independently and objectively.

Pedagogy

Based on the feasibility analysis framework of Kimbell et al. (2007), the pedagogy dimension was examined according to the extent to which assessment supported and enhanced teaching and learning. The way in which this testing method fostered English

teaching and learning is referred to as "washback" (Harmer, 2014). In this study, the washback effect mainly related to increased motivation of students to learn and perform better, and improvements in teaching speaking skills through the provision of constructive feedback and practice of self-reflection.

Students and teachers were enthusiastic about DMOVA's capacity to enhance fairness and reliability, as well as its advantages for marking and review. Such beliefs generated positive attitudes and motivation among these stakeholders. Teachers observed students were better prepared for tests, and noticed positive efforts to improve their fluency, content and delivery. This is an important finding to understand the influence of digital assessment on learning and concurs with previous studies by Green (2013); and Xie and Andrews (2013), who found the type of test had an impact on learning and preparation, i.e., a washback effect.

The results also expand upon previous research that showed some students were able to perform better when they were videoed. Teachers ascribed this to students' familiarity with the camera and sharing videos on social networks that made them feel like they were acting, especially in the group tasks. This finding casts new light on the effects of students' personal experiences with social networks and iterates the findings of De-Marcos et al. (2010), who argued that familiarity with technologies increased learner motivation, and hence, improved performance.

Teachers were more motivated to teach speaking skills after the digital assessments had been conducted accurately and fairly. Unlike Bachman and Palmer (1996), the current study did not conclude that teachers were inclined to teach to the test or change their instructions. Rather, they were motivated by this method of assessing English communication skills and wanted to teach them better.

The findings confirmed that DMOVA facilitated the provision of feedback, however, the inability to do so instantly imposed one limitation on the digital method. This was in accordance with the results of Suvorov and Hegelheimer (2014), who reported unresolved difficulties with feedback in speaking tests with computer-assisted language assessment and automatic rating of essays. Although feedback was not provided to students in real time, the teachers believed it was more detailed and comprehensive. They recognised its potential as a resource for students to reflect on their work, understand their strengths and weaknesses, and guide them towards improved performance, as asserted by Carless et al. (2011). While the washback effects that

emerged in this study were in line with many other previous findings, e.g., Green (2013); Harmer (2014); Xie and Andrews (2013), it contradicted the study of C. Chang and Lin (2019), who argued that revisions of performances could lead to stress and demotivation.

An important finding was the realisation, by both teachers and students, that they could critically reflect on their English speaking competence and assessments using the feedback and marked video recordings. A study by Stables and Kimbell (2007) indicated that digital representation provided a repository of student work and open access for student reflection, input and review by teachers. Ferrell (2012) recognised the opportunity as a source of reflection for teachers. In the current study, the student recordings served as a resource for teachers to reassess and self-reflect on their practices. DMOVA embodied this type of learning resource and repository of student oral performances for facilitating reflection and feedback, as mentioned in previous studies (Borko et al., 2008; Carless et al., 2011; C. Chang & Lin, 2019; Rosaen et al., 2008; Santagata, 2009).

The current study identified a relationship between self-reflection and validity of speaking assessments when teachers marked digitally. By reflecting on their current marking habits and how they affected accuracy, they were able to recognise aspects of the language they needed to focus on when marking (C. Chang & Lin, 2019). Being able to re-mark the recordings led them to making more accurate judgements. The anomaly of lower digital results compared to live results is broadly consistent with a study by Nakatsuhara, Inoue, and Taylor (2017), who compared IELTS examiner scores in live and recorded speaking assessments and found the video ratings lower than the live ratings. The authors concluded that teachers paid more attention to negative aspects of student performances and tended to be more critical when they marked digitally. The importance of the visual recordings was also cited by Nakatsuhara et al. (2017) as a source of information to help examiners understand students' utterances, hesitations, and pauses.

The complexities of speaking assessment were evident in this research, as there were no right or wrong answers to the test questions, making it difficult to judge which marking style was the better of the two. The findings pointed to a combination of live and digital marking as the best option for high-stakes speaking examinations, as also recommended by Nakatsuhara et al. (2017) for IELTS tests.

The student survey indicated that students were optimistic about the positive impacts of digital testing in equalising the attention paid to the four language skills in EFL assessment. It also helped to abate the issue of insufficient time for communicative practice in classrooms. H. T. Nguyen, Warren, et al. (2014) proposed implementing the digital testing method for formative assessment, with the implication that students could video their speaking performances themselves. Charman and Douglas (2006) concluded that watching their own, their friends' and sample videos for self-assessment and practice encouraged students to reflect on their speaking ability. They learn to correct their mistakes by receiving feedback from others who shared their videos, and at the same time, enhance their collaborative learning (J. Richards & Rodgers, 2014).

Technology

In the current study, the technology dimension was concerned with the compatibility of the new testing method with the existing technologies at FPT University, as clarified in the feasibility framework of Kimbell et al. (2007). Technology comprised two categories: (a) physical technologies and (b) teacher and student ICT literacy. Ease of use and potential for technical issues were also taken into consideration.

In terms of physical technologies, the Phase 1 survey results indicated that all teacher participants had laptops for teaching. Many of them used more than one technical device for their teaching and lesson planning. Ninety six percent of the 278 students possessed laptops and 76% had smartphones, which they used for study. In addition, FPT University was selected for this research because it met the technical requirements of the study. In Phase 2 the results showed that most teachers (13/18) were optimistic about the compatibility of the university's facilities with DMOVA. The results of both phases were consistent and collectively inferred that the new testing method could easily be consolidated with the available technical facilities at FPT University.

With regard to the stakeholders' ICT literacy, both research phases indicated that teachers and students were familiar with design, customisation, delivery and taking EFL computer-assisted tests. Students had not only sat computer-assisted tests for English, but other subjects too. The teachers had attended training courses on designing, customising, and delivering EFL computer-assisted tests and acquired substantial experience. The results confirmed that both teachers and students at FPT University had appropriate ICT levels for the digital testing method. Although the research was conducted at only one private university in Vietnam, these findings are still worthy of

consideration in other public universities with similar technical facilities and characteristics.

The observational data uncovered no technical issues during any of the testing sessions. The technology used for the trial were not the most recent models and teachers complained about the quality of the audio recordings on some of the iPads. To resolve the issue, they repositioned the iPad during the tests and reminded students to speak loudly. None of the teachers reported any problems with the audio quality of the videos when they marked digitally. Nevertheless, a minority of teachers were still anxious that technical faults may arise and cause delays. They were not overly confident about the potential of the digital testing method to replace teacher invigilators and thus solve the problem of EFL teacher shortages.

Teachers reported no problems with the technology because it was simple and straightforward to use. Setting up the test room and class management while video recording also created no issues. They concurred that the technology was simple and effective for English-speaking assessment and offered a variety of functions to facilitate their marking and manage the student performances. However, further training was recommended to enhance teachers' invigilation and marking skills with DMOVA.

Benefits and Limitations of Implementation

Subquestion 3: What are the benefits and limitations of digital representation of student performances for summative English speaking assessment in Vietnam?

The benefits and limitations of digital representation for summative English speaking assessment have been discussed in comparison with the current testing method. They were examined from the viewpoints of teachers and students in the context of English education at one university in Vietnam. The marking and assessment processes were taken into account to pinpoint the benefits and limitations of implementing DMOVA in real testing situations. The benefits were identified as enhanced speaking tests in relation to assessment requirements and logistics. Limitations emerged as students' nervousness in front of the camera, a lack of instant feedback, and the requirement for teachers to undergo further training.

Most teachers' perceptions of enhanced assessment were in agreement with the findings of previous studies on computer-assisted language assessment, including Barkaoui (2011), Jonsson and Svingby (2007) on fostering analytical marking; Sundqvist et al. (2018) on reducing distractions; and Kenyon and Malone (2010) on facilitating multiple

marking and review. Teachers also concurred that fairness, reliability, and validity were enhanced by the digital method, in line with the findings of Yu (2012), Kirkgoz (2011), and Costa and Kallick (2004). In contrast to a study by Pagram (2013), who concluded that teachers of Italian preferred face-to-face testing over computer-assisted testing because they found it hard to control the class and technologies, most teachers in this study preferred digital assessment.

As far as logistical advantages were concerned, the current study found most teachers liked the flexibility of digital assessment in relation to marking times and locations. The perceived benefits of marking at their convenience was consistent with the findings of Pagram (2013), who reported that the use of mobile devices contributed to the flexibility of marking assessments. In addition, the digital method reduced the manual work related to marking, recording and distributing results. These conclusions differed from Sundqvist et al. (2018), who observed a majority of respondents were not in favour of recordings because students were of the view that they took time, were administratively burdensome, and teachers did not have time to re-listen to them. Pagram (2013) also drew opposing conclusions, highlighting logistical difficulties with managing the portfolios and time for students to complete all tasks.

According to the teachers, marking group tasks digitally took longer than the face-to-face method, because they had to play back the videos multiple times. This contradicted previous research that showed recorded speaking tests supported group assessments by allowing teachers additional time for listening and consulting with colleagues (Sundqvist et al., 2018). In the current study, teachers commented that they did not have enough time to assess group tasks properly.

A further advantage of DMOVA was that marking could be done offline once the recordings were uploaded or copied from the online repository. Additionally, the recordings, embedded in the OVA App, could be saved locally and marked on the same device used to record the performance. However, uploading the recordings to the online repository and issuing different usernames and passwords required additional technical knowledge. Although digital marking did not require state-of-the-art technologies and was compatible with the facilities at FPT University, the marking platform was designed on FileMaker Pro, a software that would need to be purchased, installed, and customised by the university. The study also highlighted the need to upgrade the audio recording devices or recommend additional microphones for better quality sound recording.

Although students had overall positive perceptions of the digital testing method, many of them were evidently nervous during the tests. However, consistent with the assumptions of Yanxia (2017) and Rahimi and Zhang (2016), who also found that students were anxious about their individual English speaking proficiency and failing the test, the evidence in the current study suggested that their anxiety did not merely stem from the presence of the camera in the test room, but also other factors. This finding is consistent with Baralt and Gurzynski-Weiss (2011), who reported that face-to-face and computer-mediated communication tests had similar effects on students' states of anxiety, implying that their anxiety is likely to also originate from other sources (Huang, 2018; Yanxia, 2017). The observations confirmed that students' EFL competence was linked to their confidence. The more competent students were, the more confidently they performed, regardless of the presence of the camera. This finding was echoed by Yanxia (2017), who demonstrated that students' anxiety was predominantly caused by their low spoken English abilities and speaking techniques.

One limitation of the digital testing method was its perceived weakness in providing instant feedback as in the face-to-face method. Zhan and Wan (2016), Zhou and Yoshitomi (2019), and Phaiboonnugulkij and Prapphal (2013), all identified the positive attributes of two-way dynamic interaction and a second chance for clarification in the computer-assisted mode. Moreover, the feedback provided later was addressed in more detail and recorded as a source of study for students' reflection.

Although no technical issues were reported or observed during the speaking and marking processes, two incidents signalled the need for teacher training to avoid skipping and fast-forwarding on the OVA App. Additional features were also recommended, such as uploading recordings for use as a study source or portfolio to enhance the training content and foster best practice use of digital assessment.

Overall, the results established that once implemented, the benefits of the digital testing method outnumbered its limitations. Compared to the current face-to-face method, both teachers and students were positive and enthusiastic about the promise of logistical advantages and enhanced assessment quality. The benefits were perceived to outweigh the drawbacks, identified as student nervousness, lack of immediate feedback and teacher training requirements.

Summary

This study investigated the feasibility of implementing DMOVA in the context of a Vietnamese university. Feasibility was explored through a framework comprised of four dimensions: functionality, manageability, pedagogy, and technology. The willingness of stakeholders to use the technology, as well as the benefits and limitations of implementing it in a real testing context, were also examined.

The results of Phase 1 and Phase 2 of the study were evaluated in relation to previous studies on the same topic in the literature. Stakeholder perceptions and comparability between the test results of the digital and face-to-face marking modes were largely in line with the results presented in the literature. However, some differences were also found, leading to a new understanding of the potential of DMOVA in the context of EFL education at university level. Other findings pointed to a change in stakeholder perceptions over time and warrant further investigation in future research to cement our understanding of digital assessment.

In the current study, both teachers and students were familiar with and had experienced EFL computer-assisted assessment. In fact, this type of assessment was widely used and found to outnumber traditional paper-and-pencil tests. The teachers had attended training courses and acquired certain knowledge on using, customising, designing and delivering computer-assisted tests, in contrast to the findings of Sinwongsuwat (2012), Uzunboylu and Tuncay (2010), Hu and McGrath (2012), and Wang (2014), all from different contexts. These differing findings call for further studies on a wider scale to include multiple universities and students who are both English majors and non-majors.

In answering the research questions, the study indicates that there was indeed a lack of computer-assisted tests for speaking skills, as discovered in many other former studies, e.g., Canh, 2013; Hoang, 2010; Sinwongsuwat, 2012; and Tran, 2013. It was confirmed in both Phase 1 and Phase 2 of the study, where EFL speaking assessment was identified as the weakest aspect of English assessment. Compared to reading, writing and listening, assessment of English speaking skills is a more recent topic of research (Fulcher, 2014) and has drawn the least attention from researchers (Al Hosni, 2014). It is therefore an area worthy of further research.

The current study showed that teachers were concerned about the inability of computer-assisted speaking assessment to foster conversation and interaction and that it did not allow for instant feedback. These results were consistent with Kenyon and Malabonga

(2001), Moere (2010), Suvorov and Hegelheimer (2014), Phaiboonnugulkij and Prapphal (2013), Zhan and Wan (2016), and Zhou and Yoshitomi (2019). However, the advantages offered by DMOVA, such as fairness, reliability, consistency, validity, logistical advantages, positive pedagogical impacts and management support were recognised by most stakeholders. The technical requirements were well within the university's scope and compatible with the existing technologies. These findings were repeatedly identified and confirmed by the different data sources – survey questionnaires, interviews, observations and assessment results – confirming the hypothesis that digital testing can be feasibly implemented for EFL assessment practice at universities in Vietnam. Although feasibility has been established, future studies should take into consideration some of the limitations that were unavoidable due to time constraints and the bounds of a PhD study. These limitations are discussed further in the next chapter.

CHAPTER 7
CONCLUSIONS

This chapter presents the conclusions based on the findings that emerged from the data collected from EFL teachers and students at a university in Vietnam, using various data collection instruments throughout the two research phases of a four-year study. It adds to the existing body of knowledge on stakeholder perceptions of feasible implementation, as deduced from a comparison of the two testing methods. Results were collected from a trial of summative end-of-semester tests on English speaking performance using the digital representation method, DMOVA. The contributions of the study to the literature and the field of English speaking assessment are outlined, and the implications presented. Limitations of the study are stated and recommendations offered for future research.

Overview

There is a recognised gap in the field of EFL between what is taught and learnt and what is assessed in the English curriculum. There is also a need to include English speaking assessment in summative tests and important examinations. English speaking assessments are widely thought to motivate teachers and inspire students to learn English speaking skills. Modern technologies have been incorporated into assessment of English oral communication skills since the last decade of the 20th century, when Heaton (1990) suggested using language laboratories for speaking tests. Since then, the way English speaking is assessed has changed significantly. Moreover, there has been little research on digitisation of English speaking performance to support online marking and test administration and enhance test reliability and fairness.

This study was a response to the abovementioned issues. It investigated the feasibility of digital assessment for evaluating spoken EFL at a university in Vietnam. The research comprised two phases: Phase 1 was the preliminary stage and explored stakeholder perceptions, familiarity, and experience with computer-assisted language assessment in general and English speaking assessment in particular. The preliminary study also probed students' and teachers' willingness to participate in the digital English speaking test trial in Phase 2. The first phase involved 278 students and 17 EFL teachers from FPT University in Hanoi, Vietnam. Survey questionnaires, with both open and closed questions, were used to collect data.

Phase 2 involved 60 students with different English proficiency levels and 18 EFL teachers from the same university as in Phase 1. Both qualitative and quantitative data were collected by means of surveys, semi-structured interviews, observations and English speaking tests. Student speaking performances were marked twice, once in a traditional face-to-face interview, and again using the video presentation and OVA App. The application was customised to fit the format and purposes of the EFL speaking assessment at the university. The digital marking method offered the benefits of multiple marking and review and allowed multiple access to the online repository, as well as offline access from a mobile device. Feasibility of the implementation was analysed according to a feasibility framework (see Figure 2.7) that took into account manageability, technology, functionality and pedagogy. The benefits and limitations in the specific context of this research were also investigated.

Conclusions

The findings of the study are presented below in response to the research questions. The overarching question was: How feasible is digital representation for summative assessment of EFL speaking performance in Vietnam? The main research question was answered by three subquestions:

- What are teacher and student perceptions of computer-assisted EFL speaking assessment?
- What is the feasibility of digital representation of student performances for English speaking assessment in terms of functionality, manageability, pedagogy, and technology?
- What are the benefits and limitations of digital representation of student performances for summative English speaking assessment in Vietnam?

The key findings addressed the subquestions, discussed in relation to the literature in Chapter 6. They were categorised as stakeholders' familiarity and perceptions, feasibility dimensions, and the benefits and constraints of implementation in a Vietnamese context.

Stakeholder Perceptions and Acceptance of Digital Testing

It was evident from the results that most of the teachers and students were familiar with delivering and taking EFL computer-assisted tests. Teachers had acquired experience using, customising, designing and delivering such tests. They had also attended training courses, provided by the university, to equip them with the knowledge and skills

required for computer-assisted English tests. The survey results in both phases of the study showed that English computer-assisted tests outnumbered paper-and-pencil tests, but they were rarely used for assessing writing and speaking skills. Some teachers claimed they sometimes used computers to assist with their writing assessments, but few used them to assess speaking skills. Instead, students recorded their performance on video as a homework task.

Teachers were sceptical about the reliability of computer-assisted speaking tests, placing their trust in face-to-face interviews for authenticity and reliability. They did however recognise the drawbacks of the interview method, notably its subjectivity, the lack of test evidence, inability to review later, student distractions and fatigue after long hours of invigilation. There was some evidence in this study of a link between teachers' scepticism and their lack of experience with computer-assisted speaking tests.

All the teachers and students owned technological devices for teaching, learning and assessment. They used these devices with confidence and frequently turned to online resources for learning and teaching. The results also showed that most teachers and students demonstrated positive attitudes towards the effectiveness of computer-assisted EFL speaking assessment, perceived as enhanced transparency, flexibility and consistency.

Feasibility Dimension

To assess the implementation of DMOVA, the convergence of different data sources and comparisons of assessment results between the two marking methods were analysed according to the feasibility dimensions of functionality (Dimension A), manageability (Dimension B), pedagogy (Dimension C), and technology (Dimension D). Overall, the findings showed that both teachers and students had positive perceptions, attitudes and beliefs about using the digital assessment method for evaluating speaking skills.

The stakeholders witnessed the fairness, validity and reliability, or general functional dimension (A), enhanced by DMOVA. Most teachers concurred that it boosted fairness in EFL speaking assessment, perceived as consistency in teachers' judgements, objectivity, accuracy in marking, providing detailed feedback, and equality in the use of test time. Transparency in the assessment process, including the backup provided by the video recordings and multiple access for marking and review, were also believed to enhance objectivity, and hence, improve fairness. Perceived fairness in this study was also related to enhanced assessment validity and reliability.

The digital marking process ensured that teachers referred to predetermined criteria for their onscreen marking and steered them towards using the analytical marking method. Onscreen marking required teachers to consistently assess what they were supposed to assess, and in this way, improve the content validity of speaking assessments. Correlations between the digital and live results showed that the digital assessment method measured the same constructs as the conventional method. Any potential threats to validity were minimised by strategies, such as a confidential scoring system, to reduce teacher bias. There were no technical difficulties impacting on the assessment process, and the digital technology was deemed affordable and compatible with the university's technical facilities and the ICT background of users.

In this study, reliability was defined as accuracy and consistency of the assessment results supplied by multiple teachers marking the same performance. Consistency in teachers' judgements was one of the most important findings, crediting the video recordings and the OVA App with facilitating multiple marking, review and re-listening. Marking digitally removed the students' linguistic output from distractions and allowed teachers to mark at convenient times and locations. They were able to maintain their focus on marking student performances, because other activities associated with assessments, such as adding up results and inputting them into a computer, were all automated with the OVA App.

The results were somewhat similar and correlated for the face-to-face and digital marking methods. The live marking results correlated with the digital results for all three English levels under study. The marks awarded by teachers for the digital tests were lower than the live tests; and the individual task results, marked digitally by different teachers, were more significantly correlated than the group tasks marked the same way.

Teachers expressed positive perceptions of the manageability dimension (B), relating to setting up for tests and results management. Most agreed that the digital method successfully converted aspects of conventional EFL speaking assessments, with test evidence, results, and other logistical tasks. They found setting up for the speaking tests with DMOVA easy and encountered no technical issues during the presentations. There was strong evidence to suggest the digital testing method changed the way teachers administered their speaking assessments, and the results supported the view that DMOVA created logistical advantages.

Washback effects were the main pedagogical benefit (C) of the digital testing method. The study results showed that the digital method motivated students to prepare and perform better in their English speaking tests and encouraged teachers to provide constructive feedback and reflect on their marking. Most teachers reported that their students were better prepared for their speaking tests when they were being recorded, and some, who were familiar with technologies, performed even better than they usually did. Although not giving feedback instantly was viewed as a drawback, teachers believed they had time to provide more comprehensive comments. Teachers and students agreed that critical reflection was a distinct advantage of DMOVA.

The findings of both phases confirmed that DMOVA was well-matched with the existing technology at the university (D). The teacher and student participants were familiar with designing, customising, delivering and taking EFL computer-assisted tests, and had appropriate ICT levels. The teachers recommended an upgrade of equipment to overcome poor sound recordings. They found the test organisation and setup simple and manageable for EFL teachers, without requiring support from IT staff.

The sum of A, B, C and D led to the conclusion that all the dimensions of the feasibility framework (see Figure 3.10) were positively perceived. The most notable findings of the study were that the digital testing method enhanced assessments by enforcing review and multiple marking and facilitating results management and logistics and suited the current technology at the university and stakeholders' ICT levels. Both teachers and students expressed a preference for the digital method over the face-to-face testing approach, despite some students' nervousness in front of the camera, the lack of instant feedback, and the requirement for teachers to undergo training.

Benefits and Constraints

Enhancing the quality of assessments in relation to fairness, consistency, accuracy, validity and objectivity, was the most enduring benefit of the digital method, thought to generate positive washback effects on teaching, learning and assessment of EFL speaking skills. DMOVA changed the way speaking was assessed by allowing multiple online and offline marking. Digitisation of student performances and marking with the OVA App were widely believed to have brought about logistical advantages in relation to results submission, distribution and management; storage of test evidence; and marking confidentiality and flexibility.

However, a number of students, particularly pre-intermediates, were visibly nervous in front of the camera, raising questions about the cause of their anxiety given the results of previous studies that identified students' low English competence as the main reason for their nervousness.

The current study also raised concerns for some teacher participants, who preferred being able to provide students with instant feedback and found that digital marking took longer for group tasks. Some records went missing and overuse of the fast-forward function were reported, suggesting the need for teacher training.

Contribution

This study investigated the feasibility of digitally assessing English speaking performance at tertiary level in Vietnam. It was conducted at FPT University, which met the technical requirements of the study and included English speaking in summative end-of-semester tests. Conducting a hands-on trial using the digital testing method, DMOVA, revealed its potential as a supplementary testing method to enhance the quality of English speaking assessments.

The findings addressed a gap in our knowledge on the feasibility of using digital representation for assessing student English speaking performances. It provided a new understanding of the differences between digital and face-to-face interview assessment methods and how the process can be enhanced. From this perspective, the study contributed to improvements in the process of assessing English oral proficiency.

The research also pinpointed some problems with the current speaking assessment method and proffered suggestions on how to solve them. In addition to fostering collaborative marking and review, DMOVA addressed the enduring issues of subjectivity, and the lack of standardisation and transparency in assessments with positive results. Improved reliability, validity, impact and feasibility were additional benefits that came with modifying assessment of English oral proficiency. The OVA App changed the manner in which teachers marked student oral performances, from being a personal, individual undertaking to a public, collaborative one. The research made innovative use of onscreen marking to assess individual and group tasks; and by bringing the marking key and student performances together in one window, digitised the entire marking process.

The findings also addressed the lack of test evidence in the live method, the unavailability of recordings for review, and the scarcity of qualified English teachers to

invigilate speaking tests, while introducing concepts of peer-marking, collaborative marking and speaking portfolios. They challenged previously held views that using technologies to assess speaking skills was unauthentic and unreliable. The study confirmed that the implementation of DMOVA was feasible in tertiary EFL contexts.

Another important finding brought to light evidence that digital speaking assessment did not require advanced technologies, although training is recommended for IT staff to be able to design and customise FileMaker Pro and for teachers to smoothly manage DMOVA speaking tests. A further implication of the study was that the group task assessment needs to be revised to reduce the time and onerousness of the marking process.

Limitations of the Study

Due to the scope of a PhD study, some limitations were inevitable. First, the small sample size of the study limited generalisability of the results. In spite of this, the approach provided new insights into the feasibility of implementing a digital assessment method in a tertiary context among a specific group of real users, who enjoyed several benefits as a result. The research clearly demonstrated implementation of digital speaking assessments at university, giving rise to questions about implementation on a larger scale, in other universities, and at different school levels.

As far as the research design was concerned, the study did not include proper moderation of student results generated by either assessment method. Although moderation was undertaken by teachers when they marked live, it was as simple as the average of the overall results. The practice of class teachers invigilating their own classes in speaking tests uncovered another limitation of the study. Although this approach allowed teachers to see improvements and differences in their students' performances, it did not eliminate the risk of potential bias in their judgements.

Although adapted from the currently used marking key, some disadvantages emerged that partly affected teachers' marking, such as inadequate calibration of band scores and using the same marking key for all three different levels of proficiency. Different marking keys for different language levels should be developed to maintain consistently high accuracy and validity.

While the study generated new insights into the correlations between face-to-face and digital assessment tests, it had some limitations. First, few teachers participated in two marking rounds. Second, memorisation of their marks in the face-to-face version may

have influenced their judgement of their subsequent digital assessments. Moreover, the results may be true for one population, but not necessarily another. Given these issues, the digital method nevertheless afforded teachers opportunities to critically reflect on their marking practices, compare the face-to-face and DMOVA methods, and precisely pinpoint the pros and cons of each type of assessment.

Recommendations and Implications

In view of the limitations of the study, larger sample sizes, particularly the number of teachers marking both modes of speaking assessment, will be a valuable expansion of the findings. Similar studies in other educational contexts is also recommended, such as secondary schools and public universities, with different cohorts of participants, to explore the feasibility of DMOVA for English speaking assessment in those sectors. Determining the relationship between teacher experience and their speaking assessments was beyond the framework of the current research but will provide further insights and understanding.

Incorporating moderation in the marking process with DMOVA and further customisation of the marking keys are also recommended foci for future studies. Unlike this study that examined individual and group tasks, the inclusion of paired speaking tasks could also bring about enlightenment. Future studies could include this as a variable to further explore interactive skills and the effectiveness of digital assessment to evaluate these tasks.

Implications for Practice

The results attest to the advantages of digital assessments for evaluating university students' English speaking skills in end-of-semester tests. It could be implemented on a step-by-step basis depending on available budgets and existing technology. It is highly recommended that English tests be recorded to retain evidence of student performance for standardisation, review, and reflection. Washback effects of speaking assessment should not be underestimated, as they have an impact on developing students' communication skills and enhancing the teaching of speaking. Introducing DMOVA to EFL teachers at other universities will familiarise them with digital assessment and encourage them to reflect on their marking.

The findings show that DMOVA brings EFL speaking skills into line with other skills assessments and goes some way towards solving the current imbalance and inattention.

DMOVA is also recommended for formative assessment so that students can learn from reviewing their own performances and reflecting on teachers' feedback.

Implications for Policy

It is recommended that teachers attend training to prepare them for implementation of DMOVA and equip them with sufficient knowledge to use the equipment and method effectively. The compulsory inclusion of English speaking skills in end-of-semester tests in schools and higher educational institutions will be a catalyst for widespread change to foster improvements, regardless of whether English is a major or non-major subject. Moreover, integrated technologies should be encouraged in schools and universities for use in EFL lessons and speaking assessments.

Overall Conclusions

The findings of this study indicated that computer-assisted English assessment was popular, and in some instances, even more popular than paper-and-pencil assessments, suggesting a shift from traditional to digital assessment. Teachers and students were open and adaptable to this trend, having demonstrated their familiarity and experience with digital English assessment. The study also revealed an imbalance in the evaluation of writing and speaking skills as the two areas least often assessed digitally. The study indicated that digital representation is feasible for summative assessment of EFL speaking performance in Vietnam.

Despite evidence in the literature review of significant developments in digital assessment, including claims of accurate and reliable automated speaking assessments, actual practice has not changed much. This study identified a major gap between the development of speaking assessment and actual evaluation of this skill in schools and universities. The solution is simple and affordable and does not require state-of-the-art technologies or high levels of ICT literacy.

There were significant correlations in feasibility between the digital and face-to-face assessment methods in relation to functionality, manageability, pedagogy and technology dimensions. Participants perceived the benefits of implementing the digital method for assessing EFL speaking performance outweighed the limitations. From their perspectives, it represented a feasible improvement over the current method for assessing spoken English at tertiary level.

The data for this study were obtained from different data sources, then analysed and reviewed against the current literature to ensure the veracity of the research as a valuable source of reference for policy makers to consider changing EFL assessment schemes. It is hoped that speaking assessments will be included in EFL tests and examinations, and technologies will be introduced to enhance their quality and reliability. In the context of EFL in Vietnam, the inclusion of speaking skills in assessments could have a potentially positive impact on EFL teaching and learning, while also contributing to the goals of the National Foreign Languages Project 2020 (NFLP/ 2020 Project), the follow-up project to the NFLP/ 2020 Project and other future projects by the Ministry of Education and Training.

The benefits of using technologies in language assessment cannot be denied. It is incumbent upon policy makers, schools, universities, and teachers to adopt and implement digital assessment methods in real-life testing contexts and daily practice. Technologies are developing rapidly, but once integrated, they have the power to bring about change in every field of language assessment, including spoken assessment.

REFERENCES

Abedi, J. (2014). The use of computer technology in designing appropriate test accommodations for English language learners. *Applied Measurement in Education, 27*(4), 261-272.

Admiraal, W., Hoeksma, M., Van De Kamp, M. T., & Van Duin, G. (2011). Assessment of teacher competence using video portfolios: Reliability, construct validity, and consequential validity. *Teaching Teacher Education, 27*(6), 1019-1028.

Ahn, T. Y., & Lee, S. M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology, 47*(4), 778-786.

Airasian, P. W., & Russell, M. K. (2001). *Classroom assessment: Concepts and applications* (4th ed.). Colombus, OH: Mcgraw-Hill.

Al Hosni, S. (2014). Speaking difficulties encountered by young EFL learners. *International Journal on Studies in English Language and Literature (IJSELL), 2*(6), 22-30.

Aleksandrzak, M. (2011). Problems and challenges in teaching and learning speaking at advanced level. *Glottodidactica, 37*(1), 37-48.

Alemi, M., & Tavakoli, E. (2016). *Audiolingual method.* Paper presented at the 3rd International Conference on Applied Research in Language Studies, Iran.

Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice, 20*(1), 20-34.

Allen, A., & Joan, M. S. (2011). *Top Notch 3* (2nd ed.). New York, NY: Pearson Education ESL.

Alsied, S. M., & Pathan, M. M. (2013). The use of computer technology in EFL classroom: Advantages and implications. *International Journal of English Language & Translation Studies, 1*(1), 44-51.

Athanasou, J. A. (1997). *Introduction to educational testing*. Sydney, Australia: Social Science Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford, UK: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in Practice: Developing language assessments and justifying their use in the real world*. Oxford, UK: Oxford University Press.

Baird, J. A., Greatorex, J., & Bell, J. F. (2004). What makes marking reliable? Experiments with UK examinations. *Assessment in Education: Principles, Policy & Practice, 11*(3), 331-348.

Baleni, Z. G. (2015). Online formative assessment in higher education: Its pros and cons. *Electronic Journal of e-Learning, 13*(4), 228-236.

Baralt, M., & Gurzynski-Weiss, L. (2011). Comparing learners' state anxiety during task-based interaction in computer-mediated and face-to-face communication. *Language Teaching Research, 15*(2), 201-229.

Barkaoui, K. (2011). Effects of marking method and rater experience on ESL essay scores and rater performance. *Assessment in Education: Principles, Policy & Practice, 18*(3), 279-293.

Bashir, M., Azeem, M., & Dogar, A. H. (2011). Factor effecting students' English speaking skills. *British journal of arts and social sciences, 2*(1), 34-50.

Battaglia, M. (2008). *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage.

Bernstein, J., Moere, A. V., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing, 27*(3), 355-377.

Biggs, J. B. (2011). *Teaching for quality learning at university*. Berkshire, UK: McGraw-Hill Education.

Bloxham, S., Boyd, P., & Orr, S. (2011). Mark my words: the role of assessment criteria in UK higher education grading practices. *Studies in Higher Education, 36*(6), 655-670.

Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching Teacher Education, 24*(2), 417-436.

Brookhart, S. M., & Durkin, D. T. (2003). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education, 16*(1), 27-54.

Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing, 20*(1), 1-25.

Bull, J., & McKenna, C. (2004). *Blueprint for Computer-Assisted Assessment*. New York, NY: RoutledgeFalmer.

Burke, K. (2010). *From standards to rubrics in six steps*. Thousand Oaks, CA: Corwin Press.

Burton, R. M., & Obel, B. (2011). Computational modeling for what-is, what-might-be, and what-should-be studies—and triangulation. *Organization Science, 22*(5), 1195-1202.

Butler, Y. G. (2011). The implementation of communicative and task-based language teaching in the Asia-Pacific region. *Annual Review of Applied Linguistics, 31*, 36-57.

Campbell, A. B. (2008). *Performance enhancement of the task assessment process through the application of an electronic performance support system.* School of Education, Edith Cowan University, WorldCat.org database.

Canh, L. V. (2013). Native-English-speaking teachers' construction of professional identity in an EFL context: A case of Vietnam. *The Journal of Asia TEFL, 10*(1), 1-23.

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4), 395-407.

Carr, N. (2010). *The Shallows. What the Internet is Doing to Our Brains*. New York, NY: WW Norton.

Chalmers, D., & McAusland, W. (2014). *Computer Assisted Assessment: The Handbook for Economics Lecturers*. Glasgow, UK: Glagow Caledonian University.

Chambers, L., & Ingham, K. (2011). The BULATS online speaking test. *Research Notes, 43*(1), 21-25.

Chang, C., & Lin, H. C. K. (2019). Effects of a mobile-based peer-assessment approach on enhancing language-learners' oral proficiency. *Innovations in Education Teaching International*. Retrieved from https://srhe.tandfonline.com/doi/full/10.1080/14703297.2019.1612264

Chang, S. (2011). A contrastive study of grammar translation method and communicative approach in teaching English grammar. *English Language Teaching, 4*(2), 13-24.

Chapelle, C. A., & Douglas, D. (2006). *Assessing Language through Computer Technology*. Cambridge, UK: Cambridge University Press.

Charman, D. (1999). Issues and impacts of using computer-based assessments (CBAs) for formative assessment. In S. Brown, J. Bull, & P. Race (Eds.), *Computer-*

Assisted Assessment in Higher Education (pp. 85-94). London, UK: Kogan Page.

Chau, P. Y. (1996). An empirical investigation on factors affecting the acceptance of CASE by systems developers. *Information & Management, 30*(6), 269-280.

Chen, Z., & Goh, C. (2011). Teaching oral English in higher education: Challenges to EFL teachers. *Teaching in Higher Education, 16*(3), 333-345.

Chiedu, R. E., & Omenogor, H. D. (2014). The concept of reliability in language testing: issues and solutions. *Journal of Resourcefulness and Distinction, 8*(1), 1-9.

Chun, D., Kern, R., & Smith, B. (2016). Technology in Language Use, Language Teaching, and Language Learning. *The Modern Language Journal, 100*(S1), 64-80.

Ciula, A. (2005). Digital palaeography: using the digital representation of medieval script to support palaeographic analysis. *Digital Medievalist, 1*, 27-38.

Clark, V. L. P., & Creswell, J. W. (2008). *The mixed methods reader*. Thousand Oaks, CA: Sage.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education*. New York, NY: Routledge.

Coniam, D. (2013). The increasing acceptance of onscreen marking–The 'tablet computer'effect. *Journal of Educational Technology & Society, 16*(3), 119-129.

Cook, V. (2016). *Second language learning and language teaching*. New York, NY: Routledge.

Cooper, M. (2013). Italian Studies. In P. J. Williams & C. P. Newhouse (Eds.), *Digital Representations of Student Performance for Assessment* (pp. 125-160). Rotterdam, The Netherlands: Sense.

Costa, A., & Kallick, B. (2004). Building a self-directed community for learning: A self-assessment checklist. In *Assessment strategies for self-directed learning* (pp. 84-97). Thousand Oaks, CA: Corwin Press.

Cox, K., Imrie, B. W., & Miller, A. (2014). *Student assessment in higher education: a handbook for assessing performance*. New York, NY: Routledge.

Cox, T., & Davies, R. (2012). Using Automatic Speech Recognition Technology with Elicited Oral Response Testing. *CALICO, 29*(4), 601-618.

Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* Thousand Oaks, CA: Sage.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.

Creswell, J. W. (2014a). *A concise introduction to mixed methods research*. Thousand Oaks, CA: Sage.

Creswell, J. W. (2014b). *Research Design: Qualitative, Quantitative, & Mixed Methods Approaches*. Thousand Oaks, CA: Sage.

Crusan, D. (2012). Placement testing. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 17-25). Hoboken, NJ: Wiley/Blackwell.

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology*. New York, NY: Pearson Education ESL.

Dang, N. (Producer). (2016). Statistics of student genders of Ho Chi Minh National University. *Thanhnien*.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly, 3*(3), 319-340.

Davis, F., Bagozzi, R., & Warshaw, P. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management science, 35*(8), 982-1003.

Davis, F., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of applied social psychology, 22*(14), 1111-1132.

Davis, L. (2016). The influence of training and experience on rater performance in scoring spoken language. *Language Testing, 33*(1), 117-135.

De-Marcos, L., Hilera, J. R., Barchino, R., Jiménez, L., Martínez, J. J., Gutiérrez, J. A., . . . Otón, S. (2010). An experiment for improving students performance in secondary and tertiary education by means of m-learning auto-assessment. *Computers Education, 55*(3), 1069-1079.

De La Paz, S. (2009). Rubrics: Heuristics for developing writing strategies. *Assessment for Effective Intervention, 34*(3), 134-146.

De Vaus, D. (2013). *Surveys in social research*. New York, NY: Routledge.

Derwing, T. M., & Munro, M. J. (2009). Comprehensibility as a factor in listener interaction preferences: Implications for the workplace. *Canadian Modern Language Review, 66*(2), 181-202.

Dörnyei, Z. (2014). Motivation in second language learning. *Teaching English as a second or foreign language, 4*, 518-531.

Douglas, J. D. (1976). *Investigative social research: Individual and team field research*. Thousand Oaks, CA: Sage.

Duong, V. A., & Chua, C. S. (2016). English as a symbol of internationalization in higher education: a case study of Vietnam. *Higher Education Research Development, 35*(4), 669-683.

Edge, J. (1989). *Mistakes and Correction.* London, UK: Longman.

Ellis, R. (2010). *The Study of Second Language Acquisition.* Oxford, UK: Oxford University Press.

EPI. (2014). *Education First English Proficiency Index*. Retrieved from https://www.ef.edu/__/~/media/centralefcom/epi/downloads/full-reports/v4/ef-epi-2014-english.pdf

EPI. (2016). *Education First English Proficiency Index*. Retrieved from https://www.theewf.org/uploads/pdf/ef-epi-2016-english.pdf

EPI. (2018). *Education First English Proficiency Index*. Retrieved from https://www.ef.edu/__/~/media/centralefcom/epi/downloads/full-reports/v8/ef-epi-2018-english.pdf

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics, 5*(1), 1-4.

Facer, K., & Owen, M. (2005). The potential role of ICT in modern foreign languages learning 5-19. *NESTA Futurelab*. Retrieved from http://www.nestafuturelab.org/research/discuss/03discuss01.htm

Ferrell, G. (2012). *A View of the Assessment and Feedback Landscape: Baseline Analysis of Policy and Practice from the JISC Assessment & Feedback Programme*. Retrieved from https://www.webarchive.org.uk/wayback/archive/20140613220103/http://www.jisc.ac.uk/media/documents/programmes/elearning/AssesSment/JISCAFBaselineReportMay2012.pdf

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Thousand Oaks, CA: Sage.

Fink, A. (2012). *How to Conduct Surveys: A Step-by-Step Guide*. Thousand Oaks, CA: Sage.

Fitzpatrick, A., Davidson, D. E., Davies, G., Diakite, S., & Lund, A. (2004). Information and Communication Technologies in the Teaching and Learning of

Foreign Languages: State-of-the-Art, Needs and Perspectives. *United Nations Education, Scientific and Cultural Organisation, 1*(1), 10-26.

Flores, G. S. (2016). *Assessing English Language Learners: Theory and Practice*. New York, NY: Routledge.

Floris, F. D. (2014). Using Information and Communication Technology (ICT) to Enhance Language Teaching & Learning: An Interview With Dr. A. Gumawang Jati. *Teflin Journal, 25*(2), 139-146.

Frey, B. B., Schmitt, V. L., & Allen, J. P. (2012). Defining authentic classroom assessment. *Practical Assessment, Research, Evaluation, 17*(2), 1-18.

Fulcher, G. (2014). *Testing second language speaking*. New York, NY: Pearson Education ESL.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. New York, NY: Routledge

Fulcher, G., & Davidson, F. (2013). *The Routledge handbook of language testing*. New York, NY: Routledge.

Galaczi, E. D. (2010). Face-to-face and Computer-Based Assessment of Speaking: Challenges and Opportunities. In L. Araújo (Ed.), *Computer-Based Assessment (CBA) of Speaking Skills* (pp. 29-51). Luxembourg, Belgium: Publications Office of European Union.

Galletta, A. (2013). *Mastering the semi-structured interview and beyond: From research design to analysis and publication*. New York, NY: New York University Press.

George, D. (2011). *SPSS for windows step by step: A simple study guide and reference*. New York, NY: Pearson Education ESL.

Ghilay, Y., & Ghilay, R. (2012). Student Evaluation in Higher Education: a Comparison Between Computer Assisted Assessment and Traditional Evaluation. *i-Manager's Journal of Educational Technology, 9*(2), 8-16.

Gibbs, G. (2002). *Qualitative data analysis: Explorations with NVivo (Understanding social research)*. Buckingham, UK: Open University Press.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers educational research review, 57*(4), 2333-2351.

Gipps, C. V. (2005). What is the role for ICT-based assessment in universities? *Studies in Higher Education, 30*(2), 171-180.

Gipps, C. V., & Stobart, G. (2003). Alternative assessment. In *International handbook of educational evaluation* (pp. 549-575). New York, NY: Springer.

Gliem, J. A., & Gliem, R. R. (2003). *Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales.* Paper presented at the Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, Columbus, Ohio: Ohio State University.

Goh, C. C. M. (2007). *Teaching speaking in the language classroom*. Singapore: SEAMEO Regional Language Centre.

Green, A. (2013). Washback in language assessment. *International Journal of English Studies, 13*(2), 39-51.

Greenstein, L. (2010). *What Teachers Really Need to Know About Formative Assessment*. Alexandria, VA: ASCD Resources.

Greenstein, L. (2012). *Assessing 21st century skills: A guide to evaluating mastery and authentic learning*. Thousand Oaks, CA: Corwin Press.

Groeber, M. A., & Jackson, M. A. (2014). DREAM. 3D: a digital representation environment for the analysis of microstructure in 3D. *Integrating Materials Manufacturing Innovation, 3*(1), 56-72.

Groves, R. M. (2011). Three eras of survey research. *Public Opinion Quarterly, 75*(5), 861-871.

Hadi, S., & Zeinab, S. (2012). *Integration of ICT in language teaching: Challenges and barriers.* Paper presented at the Proceedings of the 3rd International Conference on e-Education, e-Business, e-Management and e-Learning, Singapore.

Hammond, J., & Gibbons, P. (2005). What is scaffolding?. *Teachers' voices, 8*, 8-16.

Hancock, D. R., & Algozzine, B. (2016). *Doing case study research: A practical guide for beginning researchers*. New York, NY: Teachers College Press.

Harlen, W. (2007). *Assessment of learning*. Thousand Oaks, CA: Sage.

Harmer, J. (2014). *The practice of English language teaching*. New York, NY: Pearson Education ESL.

Hart, D. (1994). *Authentic Assessment: A Handbook for Educators*. Menlo Park, CA: Addison-Wesley.

Hartle, S. (2009). What level are you? *Modern English Teacher*. Retrieved from https://www.pavpub.com/subscriptions/modern-english-teacher

Hays, P. A. (2004). Case study research. In D. Kathleen & D. L. Stephen (Eds.), *Foundations for research: Methods of inquiry in education and the social sciences* (pp. 217-234). London, UK: Lawrence Erlbaum Associates.

Heaton, J. B. (1990). *Classroom testing*. New York, NY: Longman Group.

Herbert, I. P., Joyce, J., & Hassall, T. (2014). Assessment in higher education: The potential for a community of practice to improve inter-marker reliability. *Accounting Education, 23*(6), 542-561.

Hesse-Biber, S. N. (2010). *Mixed methods research: Merging theory with practice*. New York, NY: Guilford Press.

Hewson, C. (2012). Can online course-based assessment methods be fair and equitable? Relationships between students' preferences and performance within online and offline assessments. *Journal of Computer Assisted Learning, 28*(5), 488-498.

Hiep, P. H. (2007). Communicative language teaching: Unity within diversity. *ELT Journal, 61*(3), 193-201.

Hinkel, E. (2017). Teaching Speaking in Integrated-Skills Classes. In J. I. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1-6). Hoboken, NJ: John Wiley & Sons, Inc.

Hoa, N. T. M., & Tuan, N. Q. (2007). Teaching English in primary schools in Vietnam: An overview. *Current Issues in Language Planning, 8*(2), 162-173.

Hoang, V. V. (2008). Factors affecting the quality of English education at Vietnam National University, Hanoi. *VNU Scientific Journal-Foreign Language, 24*, 22-37.

Hoang, V. V. (2010). The current situation and issues of the teaching of English in Vietnam. *Ritsumikan Studies in Language and Culture*, *22*(1), 7-18.

Holmes, N. (2015). Student perceptions of their learning and engagement in response to the use of a continuous e-assessment in an undergraduate module. *Assessment Evaluation in Higher Education, 40*(1), 1-14.

Houcine, S. (2011). *The effects of ICT on learning/teaching in a foreign language.* Paper presented at the ICT for Language Learning, Florence, Italy.

Hu, Z., & McGrath, I. (2012). Integrating ICT into College English: An implementation study of a national reform. *Education Information Technologies, 17*(2), 147-165.

Huang, H. T. D. (2018). Modeling the relationships between anxieties and performance in second/foreign language speaking assessment. *Learning Individual Differences, 63*, 44-56.

Hunter, L. (2012). Challenging the reported disadvantages of e-questionnaires and addressing methodological issues of online data collection. *Nurse researcher, 20*(1), 11-20.

Huong, T. T. (2010). Insights from Vietnam. In R. Johnstone (Ed.), *Learning through English: Policies, challenges and prospects. Insights from East Asia* (pp. 96-114). London, UK: British Council.

Igbaria, M., & Iivari, J. (1995). The effects of self-efficacy on computer usage. *Omega, 23*(6), 587-605.

Isaacs, T. (2013). International engineering graduate students' interactional patterns on a paired speaking test: Interlocutors' perspectives. In K. Mcdonough & A. Mackey (Eds.), *Second language interaction in diverse educational settings* (pp. 227-246). Amsterdam, Netherlands: John Benjamins.

Isaacs, T. (2016). Handbook of Second Language Assessment. In D. Tsagari & J. Banerjee (Eds.), (Vol. 12). Berlin, Germany: De Gruyter Mouton.

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field methods, 18*(1), 3-20.

Jackman, R. A. (2016). *Learning Strategies Employed in Communicative Language Teaching to Spur Tertiary English Majors' Communicative Competence in Real Life Situations.* I-Shou University, Taiwan, Retrieved from http://handle.ncl.edu.tw/11296/ndltd/74131860449555237302

Jamil, M., Topping, K., & Tariq, R. (2012). Perceptions of university students regarding computer assisted assessment. *TOJET, 11*(3), 267-277.

Johnson, B., & Christensen, L. (2000). *Educational research: Quantitative and qualitative approaches*. Boston, Massachusetts: Allyn & Bacon.

Johnson, B., & Turner, L. A. (2003). Data collection strategies in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social behavioral research* (pp. 297-319). Thousand Oaks, CA: Sage.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational research review, 2*(2), 130-144.

Jorgensen, D. L. (1989). *Participant observation: A methodology for human studies* (Vol. 15). Thousand Oaks: Sage.

Katz, J. (2015). A theory of qualitative methodology: The social system of analytic fieldwork. *Méthod(e)s: African Review of Social Sciences Methodology, 1*(1-2), 131-146.

Kayi, H. (2012). Teaching speaking: Activities to promote speaking in a second language. *The Internet TESL Journal, 12*(11). Retrieved from http://iteslj.org/Techniques/Kayi-TeachingSpeaking.html22.Khamkien

Ke, C., Yingwei, W., Xiaoli, H., & Yajun, Y. (2011). *Computer-assisted formative assessment in language classrooms: Focus and forms.* Paper presented at the 6th International Conference on Computer Science & Education (ICCSE), Singapore.

Kearney, J., Fletcher, M., & Bartlett, B. (2002). *Computer-based assessment: Its use and effects on student learning.* Paper presented at the Learning in Technology Education: Challenges for the 21st Century, Griffith University, Brisbane, Queenland, Australia.

Kenyon, D. M., & Malabonga, V. (2001). Comparing examinee attitudes toward computer-assisted and other proficiency assessments. *Language Learning & Technology, 5*(2), 60-83.

Kenyon, D. M., & Malone, M. (2010). Investigating examinee autonomy in a computerized test of oral proficiency. In L. Araujo (Ed.), *JRC Scientific and Technical Reports*. Luxembourg, Belgium: Publications Office of the European Union.

Khamkhien, A. (2010). Teaching English speaking and English speaking tests in the Thai context: A reflection from Thai perspective. *English Language Teaching, 3*(1), 184-190.

Khan, N., Shah, K., Farid, N., & Shah, S. (2016). Perception of High School principals' about the weak English speaking skill of teachers in district Pashawar *Asian Journal of Social Sciences & Humanities, 5*(2), 29-36.

Killen, R. (2005). *Programming and assessment for quality teaching and learning*. Melbourne, Australia: Thomson Social Science Press.

Kimbell, R. (2012a). Evolving project e-scape for national assessment. *International Journal of Technology and Design Education, 22*(2), 135-155.

Kimbell, R. (2012b). The origins and underpinning principles of e-scape. *International Journal of Technology Design Education, 22*(2), 123-134.

Kimbell, R., Wheeler, T., Miller, A., & Pollitt, A. (2007). *E-scape: E-solutions for Creative Assessment in Portfolio Environments*. London, UK: Technology Education Research Unit, Goldsmiths College.

Kirkgoz, Y. (2011). A Blended Learning Study on Implementing Video Recorded Speaking Tasks in Task-Based Classroom Instruction. *TOJET, 10*(4), 1-13.

Kirkpatrick, A. (2011). English as an Asian lingua franca and the multilingual model of ELT. *Language Teaching, 44*(2), 212-224.

Klimova, B. F. (2012). Impact of ICT on foreign language learning. *AWER Procedia Information Technology and Computer Science, 2,* 180-185.

Kozulin, A., Gindis, B., Ageyev, V. S., & Miller, S. M. (2003). *Vygotsky's educational theory in cultural context*. Cambridge, UK: Cambridge University Press.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford, UK: Pergamon Press, Inc.

Kunnan, A. J. (2013). Fairness and justice in language assessment. *The companion to language assessment, 3*, 1098-1114.

Lai, E. R., & Waltman, K. (2008). Test preparation: Examining teacher perceptions and practices. *Educational Measurement, Issues and Practice, 27*(2), 28-45.

Larson, J. W. (2000). Testing oral language skills via the computer. *Calico Journal, 18*(1), 53-66.

Laurier, E. (2010). Participant observation. In N. J. Clifford & G. Valentine (Eds.), *Key methods in geography* (pp. 133-148). Thousand Oaks, CA: Sage.

Le, H. T. (2013). ELT in Vietnam general and tertiary education from second language education perspectives. *VNU Journal of Foreign Studies, 29*(1), 65-71.

Lee, Y., Kozar, K. A., & Larsen, K. R. (2003). The technology acceptance model: Past, present, and future. *Communications of the Association for Information Systems, 12*(1), 752-780.

Li, J., & De Luca, R. (2014). Review of assessment feedback. *Studies in Higher Education, 39*(2), 378-393.

Lightbown, P. M., & Spada, N. (2013). *How Languages are Learned 4th edition-Oxford Handbooks for Language Teachers*. Oxford, UK: Oxford University Press.

Linh, V. H., Thuy, L. V., & Long, G. T. (2010). Equity and access to tertiary education: The case of Vietnam. *Working Paper 10, Development and Policies Research Center, Vietnam*.

Loumbourdi, L. (2018). Communicative Language Teaching. In J. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1-6). Hoboken, NJ: John Wiley & Son, Inc.

Luoma, S. (2004). *Assessing speaking*. Cambridge, UK: Cambridge University Press.

Lynch, T. (1997). Nudge, nudge: Teacher interventions in task-based learner talk. *ELT Journal, 51*(4), 317-325.

Mahmoud, M. S. B., Pirovano, A., & Larrieu, N. (2014). Aeronautical communication transition from analog to digital data: A network security survey. *Computer Science Review, 11*, 1-29.

Malabonga, V., Kenyon, D. M., & Carpenter, H. (2005). Self-assessment, preparation and response time on a computerized oral proficiency test. *Language Testing, 22*(1), 59-92.

Malone, D. (2012). Theories and research of second language acquisition. *Reading for day 2, Topic SLA Theories*. Retrieved from http://dl.icdst.org/pdfs/files1/cf54322e1fe40b49a0f7835cd757615f.pdf

Marangunić, N., & Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal Access in the Information Society, 14*(1), 81-95.

Margaret, E. M., & Megan, J. M. (2010). Oral Proficiency assessment: Current Approaches and Applications for Post-Secondary Foreign language Pograms. *Language and Linguistics Compass, 4(10)*, 972-986.

Maryam, K., Ahmad, H., Elham, H., & Nasrin, K. (2013). The use of ICT and technology in language teaching and learning. *Applied Science Reports, 2*(2), 46-48.

McAlpine, M. (2002). *Principles of assessment*. Glassgow, UK: University of Luton.

McGaw, B. (2006). *Assessment fit for purpose*. Paper presented at the A paper presented at the International Association for Educational Assessment, Singapore.

McIver, J., & Carmines, E. G. (1981). *Unidimensional scaling*. Thousand Oaks, CA: Sage.

McLafferty, I. (2004). Focus group interviews as a data collecting strategy. *Journal of advanced nursing, 48*(2), 187-194.

McLeod, S. A. (2018). Jean Piaget's theory of cognitive development. *Simply Psychology*, 1-9. Retrieved from https://www.simplypsychology.org/piaget.html

McNamara, T. (2000). *Language Testing*. Oxford, UK: Oxford University Press.

McNamara, T. (2011). Applied linguistics and measurement: A dialogue. *Language Testing, 28*(4), 435-440.

Mikre, F. (2010). The roles of assessment in curriculum practice and enhancement of learning. *Ethiopian Journal of Education and Sciences, 5*(2), 101-114.

Miles, M., Huberman, A. M., Huberman, M. A., & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.

Miller, D. G. (2011). *An Investigation into the feasibility of using digital representations of students' work for authentic and reliable performance assessment in applied information technology.* Edith Cowan University, Retrieved from https://ro.ecu.edu.au/theses/431/

Moere, A. V. (2010). Automated spoken language testing: Test construction and scoring model development. In L. Araújo (Ed.), *Computer-Based Assessment (CBA) of Speaking Skills* (pp. 84-99). Luxembourg, Brussels: Publications Office of the European Union.

MOET. (2008). Teaching and Learning Foreign Languages in the National Education System, Period 2008 to 2020. *1400/QĐ-TTg*. Retrieved from http://www.chinhphu.vn/portal/page/portal/chinhphu/hethongvanban?class_id=1&_page=18&mode=detail&document_id=78437

MOET. (2017). Decision of Adjustment and Supplementation of the National Foreign Languages Project 2020 for the period 2017-2025. *2080/QD-TTG*. Retrieved from http://www.ngoainguquocgia.moet.gov.vn

Morozova, Y. (2013). Methods of enhancing speaking skills of elementary level students. *Translation Journal, 17*(1), 1-24.

Morrow, K., Coombe, C., Davidson, P., O'Sullivan, B., & Stoynoff, S. (2012). Communicative language testing. In *The Cambridge guide to second language assessment*. Cambridge, Uk: Cambridge University Press.

Moskal, B. (2000). Scoring rubrics: What, When, How. *Pratical Assessment, Research and Evaluation, 7*(3), 1-5.

Mostafa, A. A. (2011). The Impact of Electronic Assessment –Driven instruction on Preservice EFL Teachers' Quality Teaching. *International Journal of Applied Educational Studies, 10*(1), 18-35.

Mullamaa, K. (2010). ICT in language learning-benefits and methodological implications. *International education studies, 3*(1), 38-44.

Nakatsuhara, F., Inoue, C., & Taylor, L. (2017). *An investigation into double-marking methods: comparing live, audio and video rating of performance on the IELTS speaking test*. Retrieved from http://hdl.handle.net/10547/622259

Nazara, S. (2011). Students' perception on EFL speaking skill development. *JET, 1*(1), 28-43.

Negoescu, A., & Boştină-Bratu, S. (2016). Teaching and learning foreign languages with ICT. *Scientific Bulletin, 21*(1), 21-27.

Newhouse, C. P. (2011). Using IT to assess IT: Towards greater authenticity in summative performance assessment. *Computers & Education, 56*(2), 388-402.

Newhouse, C. P. (2013). Applied Information Technology. In P. J. Williams & C. P. Newhouse (Eds.), *Digital Representations of Student Performance for Assessment* (pp. 49-95). Rotterdam, The Netherlands: Sense.

Newhouse, C. P., & Cooper, M. (2013). Computer-based oral exams in Italian language studies. *ReCALL, 25*(03), 321-339.

Newhouse, C. P., Williams, J., Penny, D., Pagram, J., Jones, A., Campbell, A., & Cooper, M. (2011). *Digital Forms of Assessment*. Retrieved from https://www.ecu.edu.au/schools/education/research-activity/projects/past-projects/digital-technologies/digital-forms-of-assessment

Newman, F., Couturier, L., & Scurry, J. (2010). *The Future of Higher Education: Rhetoric, Reality, and the Risks of the Market*. San Francisco, CA: Jossey-Bass.

Ngan, N. (2012). How English Has Displaced Russian and Other Foreign Languages in Vietnam since Doi Moi. *International Journal of Humanities and Social Science, 2*(23), 259-266.

Ngoc, K. M., & Iwashita, N. (2012). A comparison of learners' and teachers' attitudes toward communicative language teaching at two universities in Vietnam. *University of Sydney Papers in TESOL, 7*, 25-49.

Nguyen, H. T., Fehring, H., & Warren, W. (2014). EFL teaching and learning at a Vietnamese university: What do teachers say? *English Language Teaching, 8*(1), 31-43.

Nguyen, H. T., Warren, W., & Fehring, H. (2014). Factors Affecting English Language Teaching and Learning in Higher Education. *English Language Teaching, 7*(8), 94-105.

Nguyen, H. T. M. (2011). Primary English language education policy in Vietnam: Insights from implementation. *Current Issues in Language Planning, 12*(2), 225-249.

Nguyen, V. L. (2010). Computer mediated collaborative learning within a communicative language teaching approach: A sociocultural perspective. *The Asian EFL Journal 12*(1), 202-233.

Nguyen, V. T., & Ngo, M. K. (2015). Responses to a Language Policy: EFL Teachers' Voices. *European Journal of Social & Behavioural Sciences, 13*(2), 1830-1841.

Nicholson, S. (2015). Evaluating the TOEIC® in South Korea: Practicality, reliability and validity. *International Journal of Education, 7*(1), 221-233.

Nyroos, L., & Sandlund, E. (2014). From paper to practice: Asking and responding to a standardized question item in performance appraisal interviews. *Pragmatics Society, 5*(2), 165-190.

Orrell, J. (2005). *Assessment literacy: A precursor to improving the quality of assessment.* Paper presented at the Making a Difference: 2005 Evaluation and Assessment Conference, Sydney, NSW, Australia.

Ortega, L. (2014). *Understanding second language acquisition*. New York, NY: Routledge.

Otto, S. E. K. (2017). From Past to Present: A Hundred Years of Technology for L2 Learning. In A. C. Carol & S. Shannon (Eds.), *The Handbook of Technology and Second Language Teaching and Learning* (pp. 10-25). Oxford, UK: John Wiley & Sons, Inc.

Padurean, A., & Margan, M. (2009). Foreign language teaching via ICT. *Revista de Informatica Sociala, 7*(12), 97-101.

Pagram, J. (2013). Findings and Conclusions. In P. J. Williams & C. P. Newhouse (Eds.), *Digital representations of student performance for assessment* (pp. 197-208). Rotterdam, Germany: Sense.

Pais Marden, M., & Herrington, J. (2011). *Supporting interaction and collaboration in the language classroom through computer mediated communication.* Paper presented at the EdMedia+ Innovate Learning, Lisbon, Portugal.

Pais Marden, M., & Herrington, J. (2020). Design principles for integrating authentic activities in an online community of foreign language learners. *Educational Research, 30*(2), 635-654.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration Policy in Mental Health and Mental Health Services Research, 42*(5), 533-544.

Parker, M., & Dhanani, S. (2012). *Digital video processing for engineers: A foundation for embedded systems design*. Oxford, UK: Elsevier.

Pathan, M. M. (2012). Computer Assisted Language Testing [CALT]: Advantages, Implications and Limitations. *Research Vistas, 1*(4), 30-45.

Pearson. (2012, 02 May 2018). Into the fourth year of PTE Academic – Our story so far. Retrieved from http://pearsonpte.com/media/Documents/fourthyear.pdf

Penney, D., & Jones, A. (2013). Physical Education Studies. In P. J. Williams & C. P. Newhouse (Eds.), *Digital Representtaions of Student Performance for Assessment* (pp. 169-191). Rotterdam, The Netherlands: Sense.

Pérez-Marín, D., Pascual-Nieto, I., & Rodríguez, P. (2009). Computer-assisted assessment of free-text answers. *The Knowledge Engineering Review, 24*(4), 353-374.

Pfeffer, J. (1982). *Organizations and organization theory*. Pitman, Boston: Ballinger Publishing.

Phaiboonnugulkij, M., & Prapphal, K. (2013). Online Speaking Strategy Assessment for Improving Speaking Ability in the Area of Language for Specific Purposes: The Case of Tourism. *English Language Teaching, 6*(9), 19-29.

Piaget, J. (1976). Piaget's theory. In *Piaget and his school* (pp. 11-23). New York, NY: Springer.

Porter, P. (1986). How learners talk to each other: Input and interaction in task-centered discussions. *Talking to learn: Conversation in second language acquisition*, 200-222.

Powers, D. E. (2010). The case for a comprehensive, four-skills assessment of English-language proficiency. *R & D Connections, 14*, 1-12.

Qian, D. D. (2009). Comparing direct and semi-direct modes for speaking assessment: Affective effects on test takers. *Language Assessment Quarterly, 6*(2), 113-125.

Rahimi, M., & Zhang, L. J. (2016). The role of incidental unfocused prompts and recasts in improving English as a foreign language learners' accuracy. *The Language Learning Journal, 44*(2), 257-268.

Reynolds, C. R., Livingston, R. B., Willson, V. L., & Willson, V. (2010). *Measurement and assessment in education*. Boston, MA: Pearson Education International.

Richards, J., & Rodgers, T. (2014). *Approaches and methods in language teaching*. Cambridge, UK: Cambridge University Press.

Richards, L. (2004). *Validity and reliability? Yes! Doing it in software.* Paper presented at the Strategies Conference, University of Durham.

Rollings-Carter, F. (2010). Performance assessments versus traditional assessments. Retrieved from http://www.learnnc.org/

Rosaen, C. L., Lundeberg, M., Cooper, M., Fritzen, A., & Terpstra, M. (2008). Noticing noticing: How does investigation of video records change how teachers reflect on their experiences? *Journal of Teacher Education, 59*(4), 347-360.

Rusanganwa, J. (2013). Multimedia as a means to enhance teaching technical vocabulary to physics undergraduates in Rwanda. *English for Specific Purposes, 32*(1), 36-44.

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *34*(2), 159-179.

Salend, S. J. (2009). *Classroom testing and assessment for all students: Beyond standardization*. Thousand Oaks, CA: Corwin Press.

Salvia, J., Ysseldyke, J., & Witmer, S. (2012). *Assessment: In special and inclusive education* (12th ed.). Belmont, CA: Wadsworth Cengage Learning.

Sandelowski, M. (2000). Combining qualitative and quantitative sampling, data collection, and analysis techniques in mixed-method studies. *Research in Nursing and Health, 23*(3), 246-255.

Santagata, R. (2009). Designing video-based professional development for mathematics teachers in low-performing schools. *Journal of Teacher Education, 60*(1), 38-51.

Savignon, S. J. (2017). Communicative competence. In *The TESOL encyclopedia of English language teaching* (pp. 1-7). Hoboken, NJ: John Wiley & Sons, Inc.

Schein, E. H. (1980). *Organizational Psychology* (3rd ed.). Englewood Cliffs, New Jersey: Prentice-Hall.

Schmuller, J. (2013). *Statistical analysis with Excel for dummies*. New Jersey: John Wiley & Sons, Inc.

Seidlhofer, B. (2005). English as a lingua franca. *ELT Journal, 59*(4), 339-341.

Seidlhofer, B. (2013). *Understanding English as a lingua franca-Oxford Applied Linguistics*. Oxford, UK: Oxford University Press.

Shohamy, E. (2000). Fairness in language testing. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment: selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida* (pp. 15-19). Cambridge, UK: Cambridge University Press.

Shukla, A. A. (2018). The Enhancement of Learner Autonomy and Assessment of English Language Proficiency for young Learners through Multiple Intelligence Theory. *EPH-International Journal of Educational Research, 2*(2), 35-44.

Siccama, C. J., & Penna, S. (2008). Enhancing validity of a qualitative dissertation research study by using NVivo. *Qualitative research journal, 8*(2), 91-103.

Silverman, D. (2015). *Interpreting qualitative data*. Thousand Oaks, CA: Sage.

Simin, S., & Heidari, A. (2013). Computer-based assessment: pros and cons. *Elixir International Journal, 55*, 12732-12734.

Simpson, M., & Tuson, J. (2003). *Using Observations in Small-Scale Research: A Beginner's Guide.* Endinburgh, Scotland: Scottish Council for Research in Education.

Sinwongsuwat, K. (2012). Rethinking assessment of Thai EFL learners' speaking skills. *Language Testing in Asia, 2*(4), 75.

Snow, M. A., Kamhi-Stein, L. D., & Brinton, D. M. (2006). Teacher training for English as a lingua franca. *Annual Review of Applied Linguistics, 26*, 261-281.

Stables, K., & Kimbell, R. (2007). *Evidence through the looking glass: developing performance and assessing capability*. Paper presented at the 13th International Conference on Thinking, Norrköping, Sweden.

Stanley, G. (2013). *Language learning with technology: Ideas for integrating technology in the classroom*. Cambridge, UK: Cambridge University Press.

Stansfield, C. W., & Kenyon, D. M. (1992). Research on the comparability of the oral proficiency interview and the simulated oral proficiency interview. *System, 20*(3), 347-364.

Stigin, R., & Chapuis, J. (2012). *Introduction to student involved assessment for learning*. New York, NY: Pearson Education.

Stockwell, G. (2013). Technology and motivation in English-language teaching and learning. In E. Ushioda (Ed.), *International perspectives on motivation* (pp. 156-175). Basingstoke, Hampshire, UK: Palgrave Macmillan.

Stowell, M. (2004). Equity, justice and standards: assessment decision making in higher education. *Assessment Evaluation in Higher Education, 29*(4), 495-510.

Sundqvist, P., Wikström, P., Sandlund, E., & Nyroos, L. (2018). The teacher as examiner of L2 oral tests: A challenge to standardization. *Language Testing, 35*(2), 217-238.

Suvorov, R., & Hegelheimer, V. (2014). Computer-Assisted Language Testing. In A. J. Kunnan (Ed.), *The Companion to Language Assessment* Hoboken, NJ: Wiley-Blackwell.

Swain, M. (2005). The output hypothesis: Theory and research. In *Handbook of research in second language teaching and learning* (pp. 495-508). New York, NY: Routledge.

Tarighat, S., & Khodabakhsh, S. (2016). Mobile-assisted language assessment: Assessing speaking. *Computers in Human Behavior, 64*, 409-413.

Taylor, A. (2015). Language teaching methods: An Overview. Retrieved from https://blog.tjtaylor.net/teaching-methods/#comment-1778491883

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. *Information systems research, 6*(2), 144-176.

Thao, L., & Le, Q. (Eds.). (2011). *Technologies for enhancing pedagogy, engagement and empowerment in education: creating learning-friendly environments*. Hershey, PA: IGI Global.

Thompson, I., Buck, K., & Byrnes, H. (1989). *The ACTFL oral proficiency interview: Tester training manual*. New York, NY: American Council on the Teaching of Foreign Languages.

Thornbury, S. (2016). Communicative language teaching in theory and practice. In *The Routledge handbook of English language teaching* (pp. 242-255). New York, NY: Routledge.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. 1. *Assessment in Education, 14*(3), 281-294.

Tran, T. T. (2013). Factors affecting teaching and learning English in Vietnamese universities. *The Internet journal language, culture society, 38*(1), 138-145.

Turner, S. F., Cardinal, L. B., & Burton, R. M. (2017). Research design for mixed methods: A triangulation-based framework and roadmap. *Organizational Research Methods, 20*(2), 243-267.

Turuk, M. C. (2008). The relevance and implications of Vygotsky's sociocultural theory in the second language classroom. *Arecls, 5*(1), 244-262.

Uzunboylu, H., & Tuncay, N. (2010). Divergence of digital world of teachers. *Journal of Educational Technology Society, 13*(1), 186-194.

Van Gelder, M. M., Bretveld, R. W., & Roeleveld, N. (2010). Web-based questionnaires: the future in epidemiology? *American journal of epidemiology, 172*(11), 1292-1298.

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information systems research, 11*(4), 342-365.

Walkinshaw, I., & Duong, O. T. H. (2012). Native-and Non-Native Speaking English Teachers in Vietnam: Weighing the Benefits. *Tesl-Ej, 16*(3), 1-17.

Walkinshaw, I., & Oanh, D. H. (2014). Native and non-native English language teachers: Student perceptions in Vietnam and Japan. *Sage Open, 4*(2), 1-9.

Wang, M. J. (2014). The Current Practice of Integration of Information Communication Technology to English Teaching and the Emotions Involved in Blended Learning. *Turkish Online Journal of Educational Technology, 13*(3), 188-201.

Williams, P. J. (2013). Engineering Studies. In P. J. Williams & C. P. Newhouse (Eds.), *Digital Representations of Student Performance for Assessment* (pp. 99-122). Rotterdam, The Netherlands: Sense.

Williams, P. J., & Newhouse, C. P. (2013). *Digital representations of student performance for assessment*. Rotterdam, The Netherlands: Sense.

Winke, P. M., & Fei, F. (2008). Computer-Assisted Language Assessment. In *Encyclopedia of language and education* (pp. 1442-1453). New York, NY: Springer.

Winke, P. M., & Isbell, D. R. (2017). Computer-Assisted Language Assessment. In S. Thorne & S. May (Eds.), *Language, Education and Technology. Encyclopedia of Language and Education* (3rd ed., pp. 1-13). New York, NY: Springer.

Witt, S. M. (2012). *Automatic Error Detection in Pronunciation Training: Where we are and where we need to go.* Paper presented at the International Symposium on automatic detection on errors in pronunciation training, Stockholm, Sweden.

Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, *30*(1), 49-70.

Xiong, W., Evanini, K., Zechner, K., & Chen, L. (2013). *Automated content scoring of spoken responses containing multiple parts with factual information.* Paper presented at the Speech and Language Technology in Education, Grenoble, France.

Yanxia, Y. (2017). Test anxiety analysis of Chinese college students in computer-based spoken English test. *Journal of Educational Technology Society, 20*(2), 63-73.

Yin, R. K. (2009). *Case study research: Design and Methods*. Thousand Oaks, CA: Sage.

Young, R., & He, A. W. (1998). *Talking and testing: Discourse approaches to the assessment of oral proficiency* (Vol. 14). Amsterdam: John Benjamins.

Yu, E. (2012). Does gender, test medium, or attitude matter? Analyzing test takers' responses to technology-mediated speaking tests. *Language Testing Assessment, 1*, 1-30.

Zakrzewski, S., & Bull, J. (1998). The mass implementation and evaluation of computer-based assessments. *Assessment & evaluation in higher education, 23*(2), 141-152.

Zamorshchikova, L., Egorova, O., & Popova, M. (2011). Internet technology-based projects in learning and teaching English as a foreign language at Yakutsk State

University. *The International Review of Research in Open Distributed Learning, 12*(4), 72-76.

Zechner, K., Higgins, D., & Xi, X. (2007). *SpeechRaterTM: a construct-driven approach to scoring spontaneous non-native speech.* Paper presented at the Speech and Language Technology in Education, Farmington, PA.

Zhan, Y., & Wan, Z. H. (2016). Test takers' beliefs and experiences of a high-stakes computer-based English listening and speaking test. *RELC Journal, 47*(3), 363-376.

Zheng, X., & Davison, C. (2008). *Changing pedagogy: Analysing ELT teachers in China*. London, UK: Continuum International Publishing Group.

Zheng, Y., & Cheng, L. (2008). Test review: college English test (CET) in China. *Language Testing, 25*(3), 408-417.

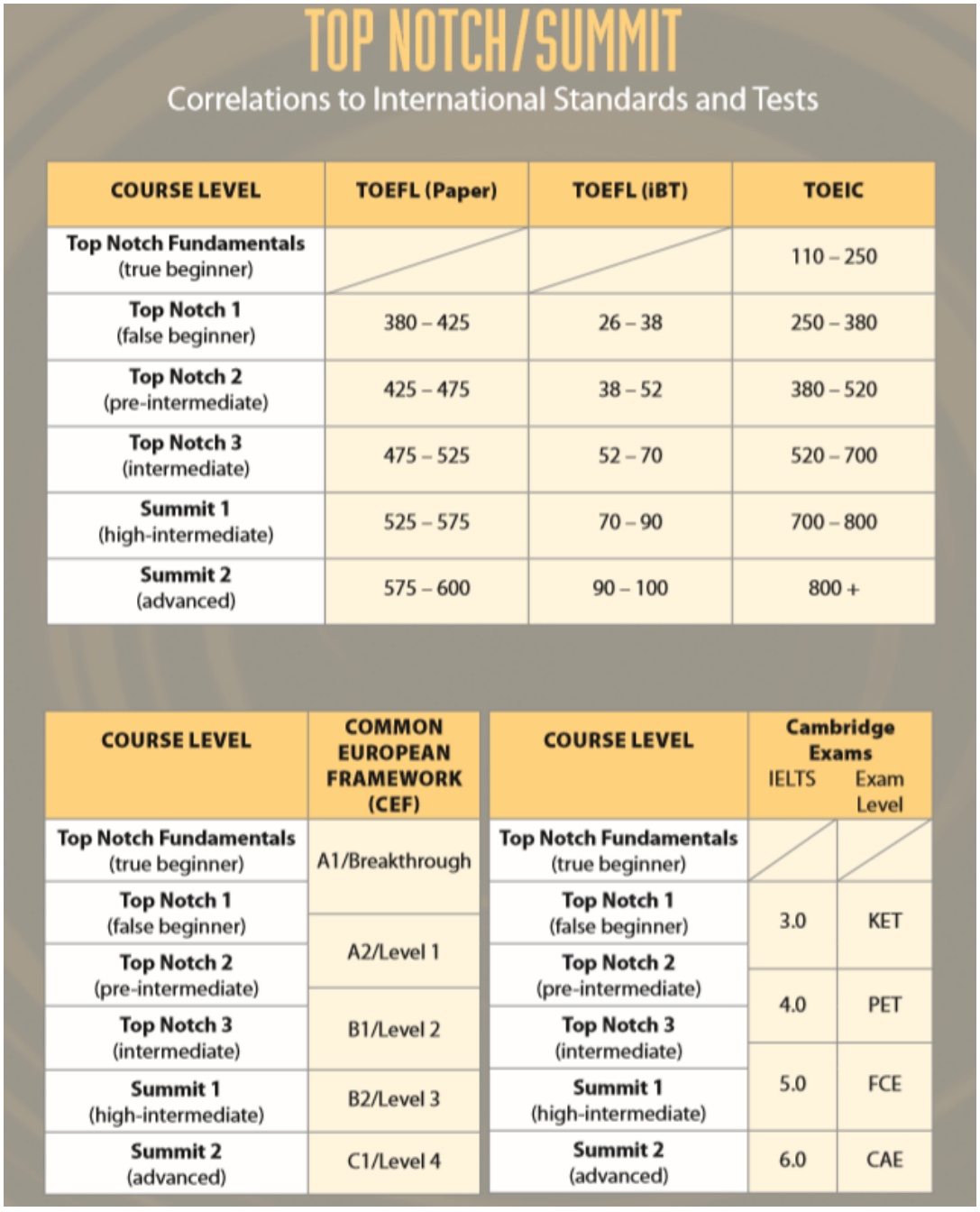
Zheng, Y., & Iseni, A. (2017). Authenticity in Language Testing. *Journal of the Association-Institute for English Language American Studies, 6*(8), 9-14.

Zhou, Y. (2015). Computer-delivered or face-to-face: effects of delivery mode on the testing of second language speaking. *Language Testing in Asia, 5*(2), 1-16.

Zhou, Y., & Yoshitomi, A. (2019). Test-taker perception of and test performance on computer-delivered speaking tests: the mediational role of test-taking motivation. *Language Testing in Asia, 9*(10), 1-19.

APPENDICES

Appendix A: Top Notch and Summit 2nd Ed. Unit-by-Unit CEF Correlations



TOP NOTCH/SUMMIT

Correlations to International Standards and Tests

COURSE LEVEL	TOEFL (Paper)	TOEFL (iBT)	TOEIC
Top Notch Fundamentals (true beginner)			110 – 250
Top Notch 1 (false beginner)	380 – 425	26 – 38	250 – 380
Top Notch 2 (pre-intermediate)	425 – 475	38 – 52	380 – 520
Top Notch 3 (intermediate)	475 – 525	52 – 70	520 – 700
Summit 1 (high-intermediate)	525 – 575	70 – 90	700 – 800
Summit 2 (advanced)	575 – 600	90 – 100	800 +

COURSE LEVEL	COMMON EUROPEAN FRAMEWORK (CEF)
Top Notch Fundamentals (true beginner)	A1/Breakthrough
Top Notch 1 (false beginner)	A2/Level 1
Top Notch 2 (pre-intermediate)	
Top Notch 3 (intermediate)	B1/Level 2
Summit 1 (high-intermediate)	B2/Level 3
Summit 2 (advanced)	C1/Level 4

COURSE LEVEL	Cambridge Exams IELTS	Exam Level
Top Notch Fundamentals (true beginner)		
Top Notch 1 (false beginner)	3.0	KET
Top Notch 2 (pre-intermediate)	4.0	PET
Top Notch 3 (intermediate)		
Summit 1 (high-intermediate)	5.0	FCE
Summit 2 (advanced)	6.0	CAE

Source: Retrieved from

http://www.pearsonlongman.com/summit2e/members/topnotch_full_course_correlation.pdf

Appendix B: Teacher interview questions, Phase Two

TEACHER INTERVIEW QUESTIONS

Semi-structured interviews

1. I would like your thoughts and feedback to be a part of my research report after you have participated in the research as assessors of students' digital representations or invigilators of the practice English speaking test, or both. Your responses will be presented anonymously by coding. Some of your responses will be directly quoted to capture your thoughts about the new English speaking assessment technique.
2. What do you think of the digital representations of students' English speaking performance for assessment?
3. To what extent do you think it was easy to use ICT to capture students' speaking performance for assessment tasks?
4. How did you feel in front of the camera? (Nervous, confident…)
5. How did the presence of the camera affect your invigilating and marking?
6. What do you think of the quality of English speaking performance produced by students, which were digitally captured?
7. What were the students' reactions to the video recording of their speaking performance?
8. What did you think about students' performance or attitude? (Were there any special cases that surprised you?)
9. What was the general feedback of students about the new English speaking assessment technique?
10. Compared to the current English speaking assessment, are the digital representations of students' English speaking performance for assessment better or worse in terms of Technical, Manageability, Pedagogic and Functional? Can you explain?
11. How much different was this to how it used to be done?
12. Did any technical problems occur within the activities?
13. How did students behave while completing the assessment tasks? (Comfort or discomfort, ease or difficulty)
14. Were there any other problems with the activities?
15. To what extent was it easy to assess students' performance digitally?
16. Do you think the results marked digitally are more reliable than the results marked in the current way? Why? Why not?
17. Did students have any problems in following the assessment tasks in front of the camera?
18. How was students' performance affected by the video recording?
19. To what extent was it easy for you to set up the camera to capture students' performance?
20. To what extent was it easy for you to keep students within the recording zone of the camera?
21. For which English level of students are the digital representations for assessment most effective, Top Notch 2, Top Notch 3, or Summit 1?
22. Which type of test are the digital representations more appropriate for summative or formative English speaking tests?
23. To what extent do you think it is feasible to implement this technique in the university context?
24. Do you think the university has appropriate technical conditions to implement this new technique for English speaking assessment?

25. Which marking method did you use when marking the digital form of students' speaking performance, Rubrics or Holistic marking? Why did you use it?
26. Do you think students prefer the new testing technique or not? Why do you think that?
27. Which English speaking assessment technique is superior, fairer, more practical in the current context of language teaching and testing in Vietnam, and more reliable, the current face-to-face live marking or digital representations of speaking performance for assessment? (Based on four dimensions)
28. Which English speaking assessment technique has better impact on English speaking teaching and learning, the current face-to-face live marking or digital representations of speaking performance for assessment?
29. Do you think that digital representations of English speaking performance for assessment help you understand how you can improve your marking? For example, you can recognise which aspects of students' performance you often miss when you mark in the current way.
30. Do you have any suggestions do you have for improving the testing technique introduced in the research?

Thank you for participating in the interview.

Appendix C: Consent Letter for Teachers

CONSENT LETTER FOR TEACHERS

DIGITAL REPRESENTATIONS FOR ASSESSMENT OF SPOKEN EFL AT UNIVERSITY LEVEL: A VIETNAMESE CASE STUDY

Thank you for your willingness to participate in the research.

The research primarily aims to investigate the reliability and the feasibility of digital representations of English speaking assessment in Vietnam. The research will involve a practice English speaking test with video recording, teacher observation and survey, and interview with a focus group of teachers. You are invited to participate in the research as an invigilator of the practice English speaking test and/or an assessor the digital representations of students' speaking performance. You can choose to be an invigilator or an assessor or both. If you choose to take part in the research, you consent to having a video taken and your voice recorded during the research.

All the information will be coded, kept confidential, and will be accessed only by the Researcher and her supervisors. Your responses may be used in a thesis or published paper. Your name and your images will not be shown in any report, thesis, or presentation of the results of this research.

The collected data will be used in my PhD studies, thesis and publications. All information will be treated confidentially and stored securely on ECU premises for ten years after the research has concluded and will then be permanently deleted.

Participation in this research is voluntary and you are free to withdraw before taking part in the practice English speaking test and there is no penalty for doing so.

If you have any questions about the research or require further information you may contact the following:

Student researcher: Thi Bich Hiep Vu. Telephone number: [redacted] or [redacted] Email: [redacted]

My supervisor: Dr Jeremy Pagram. Telephone: (+61 8) 6304 6331. Email: J.pagram@ecu.edu.au

If you have any concerns or wish to contact an independent person or an organisation about this research, you may contact:

Research Ethics Officer- Edith Cowan University. Phone: (+61 8) 6304 2170

Email: research.ethics@ecu.edu.au

I have read the Information Letter and any questions I had have been answered to my satisfaction. I freely agree to participate in the research:

I want to join as: ☐ An invigilator ☐ An assessor ☐ Both

Name: _____Signature: _____ Date: _____

Appendix D: Consent Letter for Students

CONSENT LETTER FOR STUDENTS

DIGITAL REPRESENTATIONS FOR ASSESSMENT OF SPOKEN EFL AT UNIVERSITY LEVEL: A VIETNAMESE CASE STUDY

Thank you for your willingness to participate in the research.

The research primarily aims to investigate the reliability and the feasibility of digital representations of English speaking assessment in Vietnam. The research will involve a practice English speaking test with video recording, student observation, surveys and interviews. If you choose to take part in the research, you consent to having a video taken during the practice English speaking test, and your voice audio recorded in the interviews.

All the information will be coded, kept confidential, and will be accessed only by the Researcher and her supervisors. Your responses may be used in a thesis or published papers. Your name and your images will not be shown in any report, thesis, or presentation of the results of this research. The collected data will be used in my PhD studies, thesis and publications. All information will be treated confidentially and stored securely on ECU premises for ten years after the research has been concluded and will then be permanently deleted.

Participation in this research is voluntary and you are free to withdraw before taking part in the practice English speaking test and there is no penalty for doing so.

If you have any questions about the research or require further information you may contact the following:

Student researcher: Thi Bich Hiep Vu. Telephone number: [redacted] or [redacted]. Email: [redacted]

My supervisor: Dr Jeremy Pagram. Telephone: (+61 8) 6304 6331. Email: J.pagram@ecu.edu.au

If you have any concerns or wish to contact an independent person or an organisation about this research, you may contact:

Research Ethics Officer- Edith Cowan University. Phone: (+61 8) 6304 2170

Email: research.ethics@ecu.edu.au

I have read the Information Letter and any questions I had have been answered to my satisfaction. I freely agree to participate in the research:

Name: _____Signature: _____ Date: _____

Appendix E: Teacher Observation Sheet, Phase Two

TEACHER OBSERVATION SHEET Thank you for your participation in the practice English speaking test as an invigilator – a critical part of the research. I would like to include your reactions and attitudes during the test in the research report. All the observation notes will be coded anonymously. Your name and your identity will not be identified in any reports or presentations of the research results.
CODES: 1a: Negative psychological reactions in front of the camera (nervous, worried, stressed…) 1b: Positive reactions in front of the camera (confident, engaged in the tasks, cooperative…) 2a: Gave clear instructions to students 2b: Did not give clear instructions to students. 3a: Took a long time to start. 3b: Took a short time to start. 4a: Was pleased with the test. 4b: Was dissatisfied with the test. 5a: Organised the test easily. 5b: Had difficulty in organising the test. 6a: Had problems with becoming accustomed to the presence of the camera. 6b: Did not have problems with becoming accustomed to the presence of the camera. 7a: Had some technical issues such as video recording breakdown, Wi-Fi connection, software errors. 7b: Technical issues were solved. 7c: Technical issues were not solved. 8a: Positive reactions to the new way of English speaking testing (active, relaxed, optimistic) 8b: Negative reactions to the new way of English speaking testing (annoyed, stressed, pessimistic) 9a: Took a long time to moderate students' marks in the current marking method. 9b: Took a short time to moderate students' marks in the current marking method. 10a: Positive overall reaction for the new testing technique. 10b: Negative overall reaction for the new testing technique.

Class: ….. Room: ….. University: …….. Teacher number: ……... Time period: …… to….. Date: ……..					
TEACHERS			FURTHER NOTES		
1.	Active			Video recording breakdown	
	Relaxed				
	Optimistic			Wi-Fi connection	
	Annoyed				
	Stressed			Software error	
	Pessimistic				
2.	Active			Video recording breakdown	
	Relaxed				
	Optimistic			Wi-Fi connection	
	Annoyed				
	Stressed			Software errors	
	Pessimistic				

Appendix F: Student Observation Sheet, Phase Two

STUDENT OBSERVATION SHEET Thank you for your participation in the practice English speaking test – a critical part of the research. I would like to include your reactions and attitudes during the test in the research report. All the observation notes will be coded anonymously. Your name and your identity will not be identified in any reports or presentations of the research results.
CODES: 1a: Negative psychological reactions in front of the camera (nervous, worried, stressed…) 1b: Positive reactions in front of the camera (confident, engaged in the tasks, cooperative…) 2a: Finished all the tasks. 2b: Did not finish all the tasks 3a: Took a long time to start. 3b: Took a short time to start. 4a: Was pleased with the test. 4b: Was dissatisfied with the test. 5a: Followed the instructions easily. 5b: Had difficulty in following the instructions. 6a: Had problems with becoming accustomed to the presence of the camera. 6b: Did not have problems with becoming accustomed to the presence of the camera. 7a: Had some technical issues such as video recording breakdown, Wi-Fi connection, software errors. 7b: Technical issues were solved. 7c: Technical issues were not solved. 8a: Positive reactions to the group discussion task (easy to engage in the discussion, to demonstrate performance) 8b: Negative reactions to the group discussion task (had difficulty in getting in the discussion and cooperating with one or more group members; some or one group member became too dominant) 9a: Positive reactions to the individual task (confident, demonstrated the quality in their performance). 9b: Negative reactions to the individual task (nervous, silent, hesitant) 10a: Positive overall reaction for the new testing technique. 10b: Negative overall reaction for the new testing technique.

Class: _____Room: _____University: _____Student number: _____ Time period: ____ to___ Date: _____		
STUDENTS	FURTHER NOTES	
1.	Nervous	
2.	Worried	
3.	Stressed	
4.	Confident	
5.	Engaged in the tasks	
6.	Cooperative	
7.	Video recording breakdown	
8.	Wi-Fi connection	
9.	Software errors	
10.	Easy to engage in the discussion, to demonstrate performance.	
11.	Had difficulty in getting in the discussion and cooperating with one or more group members.	
12.	Some or one group member became too dominant.	
13.	Demonstrated the quality in their performance.	
14.	Silent	
15.	Hesitant	
16.	Finished all the tasks.	
17.	Did not finish all the tasks	
18.	Took a long time to start.	
19.	Took a short time to start.	
20.	Was pleased with the test.	
21.	Was dissatisfied with the test.	
22.	Technical issues were solved.	
23.	Technical issues were not solved.	
24.	Positive overall reaction for the new testing technique.	
25.	Negative overall reaction for the new testing technique.	

Appendix G: Top Notch 2, 2nd Ed., Pearson Longman

Appendix G is not available in this version of the thesis.

The 2 images are available at: https://www.pearson.com/content/dam/one-dot-com/one-dot-com/english/TeacherResources/TopNotch/level-2-scope-sequence.pdf

Appendix H: Top Notch 3, 2nd Ed., Pearson Longman

Appendix H is not available in this version of the thesis.

The 2 images are available at:
https://pearsonerpi.com/uploads/pdf_extracts/
Top_Notch_3e_Scope_and_Sequence_Stu dent_Book_level_3_1.pdf

Appendix I: Summit 1, 2nd Ed., Pearson Longman

Appendix I is not available in this version of the thesis.

The 2 images are available at:
http://www.pearsonlongman.com/summit2e/members/level1/scope-and-sequence/scope-and-sequence.pdf

Source: Retrieved from http://www.pearsonlongman.com/summit2e/members/level1/scope-and-sequence/scope-and-sequence.pdf

Appendix J: Teacher survey questionnaire – Phase One

Q1 The integration of Information and Communication in University students' English speaking performance in Vietnam.

Thank you for your willingness to participate in the research and answer this survey which focuses on your experiences and opinions.

The survey primarily aims to investigate students and teachers' perceptions of using Information and Communication Technology in assessing students' English competence in Vietnam. If you choose to take part in the research, your responses will be sent anonymously and electronically to the researcher and may be used in a thesis or published paper. Your name will not be used at any time.

The collected data will be used in my PhD studies, thesis and publications. All information collected during the research will be treated confidentially and stored securely on ECU premises for five years after the research has concluded and will then be permanently deleted.

At the end of the survey, you will have an opportunity to register for a trial speaking test using newly developed software by entering your email address. Your email address will not be linked to your responses.

Participation in this research is voluntary and you are free to withdraw at any time before submitting the questionnaire and there is no penalty for doing so. Once you have submitted the questionnaire, collected data will be used because the data is anonymous and it is impossible to identify a participant's submission. If you have any questions about the research or require further information you may contact the following:

Student researcher: Thi Bich Hiep Vu.

Telephone number: [redacted] or [redacted]

Email: hvuthibi@our.ecu.edu.au

My supervisor: Dr Jeremy Pagram.

Telephone: (+61 8) 6304 6331.

Email: J.pagram@ecu.edu.au

If you have any concerns or wish to contact an independent person about this research, you may contact:

Research Ethics Officer- Edith Cowan University.

Phone: (+61 8) 6304 2170

Email: research.ethics@ecu.edu.au

Thank you for your time and your participation.

Q2 By clicking the next button you are giving your consent to the researcher to use your responses in the research.

Yes (1)

❍ No (2)

If No Is Selected, Then Skip To End of Survey

Q3 What is your age group?

- ❍ 18-24 years old (1)
- ❍ 25-34 years old (2)
- ❍ 35-44 years old (3)
- ❍ 45-54 years old (4)
- ❍ 55-64 years old (5)

Q4 What is your gender?

- ❍ Male (1)
- ❍ Female (2)

Q5 How long have you been teaching English?

- ❍ 0-5 years (1)
- ❍ 6-10 years (2)
- ❍ 11-15 years (3)
- ❍ 16-20 years (4)
- ❍ More than 20 years (5)

Q6 Which devices do you use to support your English teaching? (You can choose more than one answer)

- ❑ Desktop computers (1)
- ❑ Laptops (2)
- ❑ Tablets (iPad, Samsung Galaxy,...) (3)
- ❑ Smart phones (4)
- ❑ Others. Please specify (5) _____

Q7 Which websites, applications and software do you use to teach English?

- ❍ Facebook (1)
- ❍ Google Doc (2)
- ❍ Twitter (3)
- ❍ Pinterest (4)
- ❍ Gmail (5)
- ❍ Others. Please specify (6) _____

Q8 What types of English tests do you often give? (You can choose more than one answer)

- ❑ Paper-and-pencil tests (1)
- ❑ Online tests or computer-assisted tests (2)
- ❑ Oral tests (3)
- ❑ Others. Please specify (4) _____

Q9 Have you got any training on designing online tests?

- ❍ Yes. Please give the names of training courses or the tools to design online tests (1) _____
- ❍ No (2)

Q10 Do you often use English tests available online?

- ❍ Yes. Please give the names of the websites you use (1) _____
- ❍ No (2)

Q11 Do you use websites or tools to design English tests online?

- ❍ Yes. Please name the websites or tools you use to design English tests online (1) _____
- ❍ No (2)

Q12 Which English language skills do you often design online tests for? (You can choose more than one answer)

- ❑ Reading (1)
- ❑ Listening (2)
- ❑ Writing (3)
- ❑ Speaking (4)
- ❑ Others. Please specify (5) _____

Q13 Which types of English tests do you prefer?

- ❍ Paper-and-pencil tests (1)
- ❍ Computer-assisted tests or online tests (2)
- ❍ Others. Please specify (3) _____

Q14 What do you think about paper-and-pencil tests? (You can choose more than one answer)

- ❑ Reliability (1)
- ❑ Immediate feedback (2)
- ❑ Better interaction (3)
- ❑ Time-consuming (4)
- ❑ Better manageability (5)
- ❑ Authenticity (6)
- ❑ Fairness (7)
- ❑ Subjectivity (8)
- ❑ High cost (9)
- ❑ Others. Please specify (10) _____

Q15 What do you think about computer-assisted English tests or online tests? (You can choose more than one answer)

- ❑ Reliability (1)
- ❑ Immediate feedback (2)
- ❑ Better interaction (3)
- ❑ Time-consuming (4)
- ❑ Better manageability (5)
- ❑ Authenticity (6)
- ❑ Fairness (7)
- ❑ Subjectivity (8)
- ❑ High cost (9)
- ❑ Others. Please specify (10) _____

Q16 Have you ever taken a computer-assisted English speaking test with video and audio recording?

- ❍ Yes (1)
- ❍ No (2)

Q17 Have you given a computer-assisted English speaking test with video and audio recording to your students?

❍ Yes (1)
❍ No (2)

Q18 What types of English speaking tests do you often give to your students?

❍ Face-to-face interviews (1)
❍ Computer-assisted English speaking tests with video and audio recording (2)
❍ Others. Please specify (3) _____

Q19 What do you think about current face-to-face interviews in English speaking tests? (You can choose more than one answer)

❑ Others. Please specify (10) _____

Q16 Have you ever taken a computer-assisted English speaking test with video and audio recording?

❍ Yes (1)
❍ No (2)

Q17 Have you given a computer-assisted English speaking test with video and audio recording to your students?

❍ Yes (1)
❍ No (2)

Q18 What types of English speaking tests do you often give to your students?

❍ Face-to-face interviews (1)
❍ Computer-assisted English speaking tests with video and audio recording (2)
❍ Others. Please specify (3) _____

Q19 What do you think about current face-to-face interviews in English speaking tests? (You can choose more than one answer)

❑ Reliability (1)
❑ Immediate feedback (2)
❑ Better interaction (3)
❑ Time-consuming (4)
❑ Better manageability (5)
❑ Authenticity (6)
❑ Fairness (7)
❑ Subjectivity (8)
❑ High cost (9)
❑ Recording for later review (10)
❑ Others. Please specify (11) _____

Q20 What do you think about computer-assisted English speaking tests with video and audio recording? (You can choose more than one answer)

❑ Reliability (1)
❑ Immediate feedback (2)
❑ Better interaction (3)
❑ Time-consuming (4)
❑ Better manageability (5)
❑ Authenticity (6)
❑ Fairness (7)
❑ Subjectivity (8)

- ❑ High cost (9)
- ❑ Recording for later review (10)
- ❑ Others. Please specify (11) _____

Q21 Would you like to use computer-assisted English speaking tests instead of current face-to-face interviews?

- ❍ Yes (1)
- ❍ No (2)
- ❍ Maybe (3)
- ❍ Please give reasons (4) _____

Q22 Would you like to use a sample computer-assisted English speaking test as a practice test for your students?

- ❍ Yes. (Please give your email address) (1) _____
- ❍ No (2)
- ❍ I'm not sure. I want you to contact me later. (Please give your email address) (3) _____

Appendix K: Student survey questionnaire – Phase One

Q1 The integration of Information and Communication in University students' English speaking performance in Vietnam.

Thank you for your willingness to participate in the research and answer this survey which focuses on your experiences and opinions.

The survey primarily aims to investigate students and teachers' perceptions of using Information and Communication Technology in assessing students' English competence in Vietnam. If you choose to take part in the research, your responses will be sent anonymously and electronically to the researcher and may be used in a thesis or published paper. Your name will not be used at any time.

The collected data will be used in my PhD studies, thesis and publications. All information will be treated confidentially and stored securely on ECU premises for five years after the research has concluded and will then be permanently deleted.

At the end of the survey, you will have an opportunity to register for a trial speaking test using newly developed software by entering your email address. Your email address will not be linked to your responses.

Participation in this research is voluntary and you are free to withdraw at any time before submitting the questionnaire and there is no penalty for doing so. Once you have submitted the questionnaire, collected data will be used because the data is anonymous and it is impossible to identify a participant's submission. If you have any questions about the research or require further information you may contact the following:

Student researcher: Thi Bich Hiep Vu.

Telephone number: [redacted] or [redacted]

Email: [redacted]

My supervisor: Dr Jeremy Pagram.

Telephone: (+61 8) 6304 6331.

Email: J.pagram@ecu.edu.au

If you have any concerns or wish to contact an independent person about this research, you may contact:

Research Ethics Officer- Edith Cowan University.

Phone: (+61 8) 6304 2170

Email: research.ethics@ecu.edu.au

Thank you for your time and your participation.

Q2 By clicking the next button you are giving your consent to the researcher to use your responses in the research.

Yes (1)

No (2)

If No Is Selected, Then Skip To End of Survey

Q3 What is your year of birth?

_____ 1960 (1)

_____ 1961 (2)

_____ 1962 (3)

_____ 1963 (4)

_____ 1964 (5)

_____ 1965 (6)

_____ 1966 (7)

_____ 1967 (8)

_____ 1968 (9)

_____ 1969 (10)

_____ 1970 (11)

_____ 1971 (12)

_____ 1972 (13)

_____ 1973 (14)

_____ 1974 (15)

_____ 1975 (16)

_____ 1976 (17)

_____ 1977 (18)

_____ 1978 (19)

_____ 1979 (20)

_____ 1980 (21)

_____ 1981 (22)

_____ 1982 (23)

_____ 1983 (24)

_____ 1984 (25)

_____ 1985 (26)

_____ 1986 (27)

_____ 1987 (28)

_____ 1988 (29)

_____ 1989 (30)

_____ 1990 (31)

_____ 1991 (32)

_____ 1992 (33)

_____ 1993 (34)

_____ 1994 (35)

_____ 1995 (36)

_____ 1996 (37)

_____ 1997 (38)

_____ 1998 (39)

_____ 1999 (40)

_____ 2000 (41)

_____ Not applicable (42)

Q4 Are you male or female?

Male (1) Female (2)

Q5 How long have you been learning English?

_____ 1 year (1)

_____ 2 years (2)

_____ 3 years (3)

_____ 4 years (4)

_____ 5 years (5)

_____ 6 years (6)

_____ 7 years (7)

_____ 8 years (8)

_____ 9 years (9)

_____ 10 years (10)

_____ 11 years (11)

_____ 12 years (12)

_____ 13 years (13)

_____ 14 years (14)

_____ 15 years (15)

_____ Not applicable (16)

Q6 What level of English are you learning?

Beginner (1)

Elementary (2)

Pre-Intermediate (3)

Intermediate (4)

Upper-Intermediate (5)

Pre-Advanced (6)

Advanced (7)

Not applicable (8)

Q7 Do you have English tests at the end of semesters?

Yes (1) No (2)

Q8 What types of English tests do you often have? (You can choose more than one answer)

Paper-and-pencil tests (1)

Computer- assisted tests (2)

Oral tests (3)

Others. (Please specify) (4) _____

Q9 Which types of English tests do you prefer?

Paper-and-pencil tests. Can you give the reasons why? (1) _____

Computer-assisted tests. Can you give the reasons why? (2) _____

Oral tests. Can you give the reasons why? (3) _____

Others. (Please specify) (4) _____

Q10 Which English skills are you having online tests or computer-assisted tests for? (You can choose more than one answer)

Reading (1)

Listening (2)

Writing (3)

Speaking (4)

Q11 Which online tests would you prefer? (You can choose more than one answer)

Reading (1)

Writing (2)

Listening (3)

Speaking (4)

Q12 Do you learn English speaking skills in your English lesson?

Yes (1) No (2)

I do not know. (3)

Q13 Do you have an English speaking test at the end of each semester?

Yes (1)

No (2)

If No Is Selected, Then Skip To What types of digital equipment do you...

Q14 What kind of English speaking tests do you often have? (You can choose more than one answer)

Face-to face teacher and student interviews (1)

Group discussion with teacher's observation and judgment (2)

Both interviews and group discussion (3)

Speaking to a computer with audio and video recording (4)

Face-to-face interviews with audio recording (5)

Others. (Please specify) (6) _____

Q15 What do you think about face-to-face interviews in English speaking tests? (You can choose more than one answer)

Better interaction (1)

Immediate feedback (2)

Authenticity (3)

Records for later review (4)

Time-consuming (5)

Stress (6)

Nervousness (7)

Unreliability (8)

Unfairness (9)

Subjectivity (10)

Others. (Please specify) (11) _____

Q16 Have you ever taken an English speaking test in a computer-assisted format?

Yes (1)

No (2)

Q17 Do you think computer-assisted English speaking tests with audio and video recording are a good idea?

Yes (1)

No (2)

Others. (Please specify) (11) _____

Q18 If you have a choice, which type of English speaking test would you like to take?

Current face-to-face interviews (1)

Computer-assisted English speaking tests (2)

Others. (Please specify) (3) _____

Q19 Which devices do you use to support your English study? (You can choose more than one answer)

Personal computers (1)

Laptops (2)

Smart phones (3)

Tablets (iPhone, Samsung galaxy Tab, ...) (4)

Public computers (5)

Others. (Please specify) (6) _____

Q20 How often do you use digital equipment to study English?

Every day (1)

Three or more times a week (2)

Once a week (3)

Rarely (4)

Never (5)

Others. (Please specify) (6) _____

Q21 Can you use the following applications and websites to study English? (You can choose more than one answer)

English language learning websites. If Yes, can you name some of them? (1) _____

Facebook (2)

Google Doc (3)

Twitter (4)

Pinterest (5)

WhatsApp (6)

LinkedIn (7)

Others. (Please specify) (8) _____

Q22 Would you like to join a trial computer-assisted English speaking test without teachers' observation?

Yes. Please give your email address (1) _____

No (2)

I'm not sure. If you want to have later contact, please give your email address (3) _____

Appendix L: Marking key for group discussions and individual responses

Criteria	Type		Mark	0	1	2	3	4
Fluency	Group	1	3	No communication possible.	Pauses are frequent and lengthy. Uses mainly simple sentences. Gives only simple and short responses and is frequently unable to convey basic message.	Is able to speak at length, though sometimes loses coherence due to occasional repetition, self-correction or hesitation. Is able to use a range of connectives and discourse markers but not always appropriately	Speaks fluently with little repetition or self-correction. Any hesitation is idea-related rather than to find words or grammar. Speaks coherently with suitable cohesive features. Develops topics fully and appropriately	x
Pronunciation	Group	2	2	No communication possible.	Uses a limited range of pronunciation features correctly. Mispronunciations are frequent and cause some difficulty for the listener.	Uses a wide range of pronunciation features correctly. Maintains flexible use of features, with few occasional	X	x

						lapses. Is easy to understand throughout. Native language accent has minimal interference on intelligibility.		
Accuracy	Group	3	3	No communication possible.	Attempts to use basic sentence forms with little success, or relies on memorised utterances. Makes numerous errors.	Uses a mix of simple and complex structures, but with limited flexibility. May make frequent mistakes with complex structures though these rarely cause comprehension problems.	Uses a full range of structures naturally and appropriately. Produces consistently accurate structures.	X
Lang & Expression	Group	4	4	No communication possible.	Only produces isolated words or memorised utterances.	Is able to discuss familiar topics but can only convey little on unfamiliar topics and makes frequent errors in	Uses vocabulary flexibly to discuss a variety of topics, including some less common words and idioms. Has some choices of style and	Uses vocabulary flexibly and appropriately in all topics.

						word choice. Rarely paraphrases.	collocation, but they are inappropriate. Uses paraphrase effectively.	Uses idiomatic language naturally and accurately.
Total			12					
Fluency	Ind	1	2	No communication possible.	Pauses are frequent and lengthy. Uses mainly simple sentences. Gives only simple and short responses and is frequently unable to convey basic message.	Speaks fluently with little repetition or self-correction. Any hesitation is idea-related rather than to find words or grammar. Speaks coherently with suitable cohesive features. Develops topics fully and appropriately	x	x

Pronunciation	Ind	2	2	no communication possible	Uses a limited range of pronunciation features correctly. Mispronunciations are frequent and cause some difficulty for the listener.	Uses a wide range of pronunciation features correctly. Maintains flexible use of features, with few occasional lapses. Is easy to understand throughout. Native language accent has minimal interference on intelligibility.	x	x
Lang & Expression	Ind	3	2	No communication possible.	Is able to discuss familiar topics but can only convey little on unfamiliar topics and makes frequent errors in word choice. Rarely paraphrases.	Uses vocabulary flexibly and appropriately in all topics. Uses idiomatic language naturally and accurately.	x	X

Content	Ind	4	2	No communication possible.	Can talk about the topic but simply with little understanding. Content is limited and not always relevant.	Expresses a large number of relevant ideas about the topic with deep understanding and details.	x		x
Total			8						
Total			20						

Appendix M: Marking Paper Sheet

ENGLISH SPEAKING TEST – MARKING PAPER

Level: Date: _____ Room: ____

No	Student ID	NAME	Group-work task: 12 marks					Individual task: 8 marks					Total	Student's Signature
			Fluency	Pronun-ciation	Accu-racy	Lang & Expression	Sub-Total	Fluency	Pronun-ciation	Lang & Expression	Content	Sub-Total		
			3	2	3	4	12	2	2	2	2	8	20	
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														
11														
12														
13														
14														
15														
16														
17														
18														
19														
20														
21														
22														
23														
24														
25														
									Invigilator 1			Invigilator 2		

Appendix N: Teacher survey questionnaire – Phase Two

PhD - Teacher survey - 2018

Q1 Thank you very much for participating in our survey. We appreciate your feedback. In this survey, the term: "Digital representations of students' EFL speaking performance for assessment" is basically equal to "The video recording of EFL speaking performance for assessment".

Q2 Your year of birth:

_____

Q3 Your gender:

Male (1)

Female (2)

Transgender (3)

Others (4) _____

Q4 How long have you been teaching English? (How many years?)

_____

Q5 The integration of ICT in EFL (English as a Foreign Language) assessment.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
I have used, adapted, designed and given students EFL exams/tests using ICT before. (1)					
I am used to using, adapting, designing and giving students EFL exams/tests using ICT. (2)					
I often use, adapt,design and give students EFL Vocabulary exams/tests using ICT. (3)					
I often use, adapt,design and give students EFL Grammar exams/tests using ICT. (4)					
I often use, adapt, design and give students EFL Reading exams/tests using ICT. (5)					
I often use, adapt, design and give students EFL Writing exams/tests using ICT. (6)					
I often use, adapt, design and					

give students EFL Listening exams/tests using ICT. (7)	
I often use, adapt, design and give students EFL Speaking exams/tests using ICT. (8)	
I have ever recorded videos of my students' English speaking for assessment. (9)	
I have ever assigned my students tasks of videoing their English speaking for further practice at home. (10)	
I have ever assigned my students tasks of videoing their English speaking for assessment. (11)	
I like using, adapting, designing and giving students EFL exams/tests using ICT. (12)	
EFL exams/tests using ICT outnumber paper-based exams/tests at my university. (13)	

Q6 Benefits of digital representations of EFL speaking performance for assessment.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
Video recording of my students' EFL speaking is a good way to reflect their English speaking performance for assessment tasks. (1)					
Videos of my students' English speaking performance for assessment tasks would be backup for me to review their performance later. (2)					
Videos of my students' English speaking performance for assessment tasks would provide evidence of their speaking performance and their exam attendance. (3)					
Digital representations of EFL speaking performance for assessment would backup records of my students' performance, which is similar to other language skill assessment. (4)					
Videos of my English speaking performance for assessment tasks would better show me their strengths and weaknesses that I can not fully recognise when I do the marking in the current way. (5)					
Digital representations of English speaking					

performance for assessment are useful for explaining the process of my students' performance. (6)	
Digital representations of English speaking performance for assessment may enhance EFL speaking assessment quality. (7)	
Thanks to videoing of my students' EFL speaking performance, my students focus more not only on their content and fluency but also on their speaking manners. (8)	
I see my students are usually better-prepared for their EFL speaking performance when their performance is videoed. (9)	
Digital representations of EFL speaking for assessment may help English speaking assessment have equal role as the other English skill assessment. (10)	
It was easy to manage the technologies and the test at the same time. (11)	
One invigilator can manage the technologies and the test at the same time. (12)	
University's available facilities can be feasible for digital representations of EFL speaking for assessment. (13)	
Digital representations of EFL speaking for assessment do not require English teachers to be invigilators. (14)	
Overall, digital representations of English speaking performance for assessment are good for English speaking assessment. (15)	
Overall, it is better doing the English speaking assessment tasks using digital representations than doing those in the current way. (16)	

Q7 Teachers' interest in digital representations of EFL speaking performance for assessment.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
It's a good idea to have my students' EFL speaking performance video recorded. (1)					
Using digital representations of English speaking performance for assessment may enhance my EFL speaking skill teaching. (2)					
Using digital representations of English speaking performance for assessment is a good way to support EFL speaking assessment. (3)					
I am positive about the reliability and					

feasibility of using digital representations of English speaking performance for assessment. (4)	
I believe that digital representations of English speaking performance for assessment cold be a more reliable way of doing assessment. (5)	
I enjoyed using digital representations of English speaking performance for assessment. (6)	

Q8 Teachers' perspectives of how digital representations of EFL speaking performance is marked

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
It is a real difference: I can watch and re-watch the videos, listen and re-listen to students' performance to give them the best feedback and the most accurate results. (1)					
Videos of my students' English speaking help me assess their English speaking skills more equitably and comprehensively. (2)					
Videos of my English speaking performance for assessment tasks help me review students' performance later. (3)					
It is fairer to mark digital representations compared to live marking . (4)					
It is more reliable to mark digital representations compared to live marking . (5)					
It is easy to mark digital representations of students' EFL speaking performance. (6)					
My feedback would be recorded in the Marking Tool and help my students understand what aspects they should improve in their next performance. (7)					
Digital representations of EFL speaking performance allows peer-reviewing and multi-marking. (8)					
Digital representations of EFL speaking performance for assessment help me understand how I can improve my marking. (9)					
The Marking Tool was easy for me to mark and export the results. (10)					
The Marking Tool was innovative, user-friendly, and supportive. (11)					
It is easy to recognise individual in the group-work task. (12)					
It is easy to mark group-work tasks. (13)					

It is easy to mark individual tasks. (14)	
It is easy to input feedback in the Marking key. (15)	
I can do the marking at my convenient time. (16)	

Q9 Teachers' comments on the quality of videos

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
The quality of the videos is good. (1)					
The image quality of videos is good. (2)					
The sound quality of videos is good. (3)					
The videos truly capture and reflect students' performance. (4)					
It is easy to access to the Marking Tool to mark videos of students' EFL speaking performance. (5)					
The videos can be run on any digital devices, such as iPad, laptops, smart phones, and iMac. (6)					

Q10 Teachers' interest of different aspects of the new digital EFL speaking assessment.

	Very dissatisfied (1)	Dissatisfied (2)	Neutral (3)	Satisfied (4)	Very satisfied (5)
Marking of students' speaking performance. (1)					
The reliability of the test results. (2)					
The validity of the assessment. (3)					
The economical features of applying this testing method. (4)					
The application of new technology in the exam/test. (5)					
The pedagogical effects (The testing method may support and enhance EFL speaking teaching and learning). (6)					
The backup of students' EFL speaking performance. (7)					
Ease of the practice of this testing method. (8)					
The flexibility of this testing method. (9)					
The effectiveness of this testing method in assessing EFL speaking skills. (10)					
The feasibility of this testing method with University available resources. (11)					

Q11 Teachers' interest of different aspects of current speaking assessment, which is being used now at your university.

	Very dissatisfied (1)	Dissatisfied (2)	Neutral (3)	Satisfied (4)	Very satisfied (5)
Management of the exam/test. (1)					
Marking of students' speaking performance.(2)					
The reliability of the test results. (3)					
The validity of the assessment. (4)					
The economical features of applying this testing method. (5)					
The application of new technology in the exam/test. (6)					
The pedagogical effects (The testing method may support and enhance EFL speaking teaching and learning). (7)					
Time required to set up and finish the test. (8)					
The organisation of the exam/test. (9)					
The backup of students' EFL speaking performance. (10)					
Ease of the practice of this testing method. (11)					
The flexibility of this testing method. (12)					
The effectiveness of this testing method in assessing EFL speaking skills. (13)					
The feasibility of this testing method with University available resources. (14)					

Q12 Two things that I like best about digital representations of EFL speaking for assessment.

_____

Q13 Two things that I do not like about digital representations of EFL speaking for assessment.

_____

Q14 Which assessment task is more effective using digital representations? Why?

The group-work task. (1) _____

The individual task. (2) _____

Both of them. (3) _____

None of them. (4) _____

Q15 When you do the marking in the current way, what marking method do you use?

I use analytical marking method. (1)

I use holistic marking method. (2)

I often switch between the two methods. (3)

Q16 When you did the marking digitally, what marking method did you use?

I used analytical marking method. (1)

I used holistic marking method. (2)

I often switched between the two methods. (3)

Q17 Have you got any suggestions for improving the Marking Tool introduced in the research? What are they?

Yes. (1) _____

No. (2) _____

Q18 Were there any technical problems with doing the activities? What were they?

Yes. (1) _____

No. (2) _____

Q19 Were there other problems with the activities? What were they?

Yes. (1) _____

No. (2) _____

Q20 Have you got any suggestions for improving the use of digital representations of EFL speaking for assessment? What are they?

Yes. (1) _____

No. (2) _____

Q21 Which of the following activities would the digital representations of students' EFL speaking performance be more effective? (You can choose more than one answer).

Reviewing students' performance after the exam. (1)

Recording the evidence of students' performance. (2)

EFL speaking summative tests. (3)

EFL speaking formative tests. (4)

Student's homework tasks. (5)

Supporting the current EFL speaking assessment methods. (6)

High-stakes EFL speaking assessment, such as University entrance exams. (7)

Can you suggest other usage of digital representations in EFL assessment? (8)

_____

Appendix O: Student Survey Questionnaire – Phase Two

PhD - Student survey - 2018

Q1 Thank you very much for participating in our survey. We appreciate your feedback. In this survey, the term: "Digital representations of students' EFL speaking performance for assessment" is basically equal to "The video recording of EFL speaking performance for assessment".

Q2 Your year of birth:

_____

Q3 Your gender:

Male (1)

Female (2)

Transgender.(3)

Others (4) _____

Q4 How long have you been learning English? (How many years?)

Q5 The integration of ICT in the examinations in general.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
I have taken an examination or a test using ICT before. (1)					
I am used to taking exams/tests using ICT. (2)					
I like taking exams/tests using ICT. (3)					
Exams/tests using ICT outnumber paper-based exams/tests at my university. (4)					

Q6 The integration of ICT in the English as a foreign language examinations/tests.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
I have taken an EFL examination or a test using ICT before. (1)					
I am used to taking EFL exams/tests using ICT. (2)					
I often take EFL Reading exams/tests using ICT. (3)					

I often take EFL Writing exams/tests using ICT. (4)
I often take EFL Listening exams/tests using ICT. (5)
I often take EFL Speaking exams/tests using ICT. (6)
I have ever recorded videos of my English speaking for practice. (7)
I have ever recorded videos of my English speaking for assessment. (8)
I often take EFL Vocabulary exams/tests using ICT. (9)
I often take EFL Grammar exams/tests using ICT. (10)
I like taking EFL exams/tests using ICT. (11)
EFL exams/tests using ICT outnumber paper-based exams/tests at my university. (12)

Q7 Benefits of digital representations of English speaking performance for assessment.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
Video recording of my English speaking is a good way to reflect my English speaking performance. (1)					
Videos of my English speaking performance for assessment tasks would be samples for me to review my performance. (2)					
Videos of my English speaking performance for assessment tasks would provide evidence of my speaking performance and my exam attendance. (3)					
Digital representations of EFL speaking performance for assessment would provide records of my performance, which is similar to other language skill assessment. (4)					
Videos of my English speaking performance for assessment tasks would show me my strengths and weaknesses that I can not recognise myself without videos. (5)					
I am usually better-prepared for my EFL speaking performance because it would be recorded assessment. (6)					

Thanks to videoing of my EFL speaking performance assessment, I focus more on learning EFL speaking skills; therefore, my EFL speaking become better. (7)	
Thanks to videoing of my EFL speaking performance, I focus more not only my content and fluency but also on my speaking manners. (8)	
Digital representations of English speaking performance for assessment are useful for explaining the process of my performance. (9)	
Digital representations of English speaking performance for assessment may enhance my assessment results. (10)	
Overall, digital representations of English speaking performance for assessment are good for English speaking assessment. (11)	
Overall, it is better doing the English speaking assessment tasks using digital representations than doing those in the current way. (12)	

Q8 Students' interest in digital representations of EFL speaking performance for assessment.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
I am confident in front of the camera. (1)					
I feel OK about being videoed in my EFL speaking test. (2)					
I like to have my performance video recorded. (3)					
Using digital representations of English speaking performance for assessment may enhance my performance. (4)					
Using digital representations of English speaking performance for assessment is a good way to support EFL speaking assessment. (5)					
I am positive about the reliability and feasibility of using digital representations of English speaking performance for assessment. (6)					
I believe that digital representations of English speaking performance for assessment cold be a more reliable way of doing assessment. (7)					
I enjoyed using digital representations of					

English speaking performance for assessment. (8)	

Q9 Students' perspectives of how digital representations of EFL speaking performance would be assessed.

	Strongly disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly agree (5)
It is a real difference: my teachers can watch and re-watch my video, listen and re-listen to my performance to give me the best feedback and accurate results. (1)					
Videos of my English speaking help my teachers assess my English speaking skills more equitably and comprehensively. (2)					
Videos of my English speaking performance for assessment tasks help teachers review my performance later. (3)					
The assessment is fairer compared to the current assessment. (4)					
The assessment is more reliable compared to the current assessment. (5)					
Teachers' feedback would be recorded and help me understand how I can improve my performance. (6)					
I can share videos of my EFL speaking with friends and get their comments. (7)					

Q10 Students' interest of digital representation test procedure.

	Very dissatisfied (1)	Somewhat dissatisfied (2)	Neutral (3)	Somewhat satisfied (4)	Very satisfied (5)
The technologies used in the test room. (1)					
The position of the camera. (2)					
The waiting time before the test. (3)					
The size of the group (4 students). (4)					
The test room. (5)					
The individual speaking					

task. (6)	
The group-work speaking task. (7)	
The time needed to finish the test. (8)	
The process of videoing the test. (9)	

Q11 Students' interest of different aspects of the current EFL speaking assessment.

	Very dissatisfied (1)	Dissatisfied (2)	Neutral (3)	Satisfied (4)	Very satisfied (5)
Management of the exam/test. (1)					
Marking of students' speaking performance. (2)					
The reliability of the test results. (3)					
The validity of the assessment. (4)					
The economical features of applying this testing method. (5)					
The application of new technology in the exam/test. (6)					
The pedagogical effects (The testing method may support and enhance EFL speaking teaching and learning). (7)					
Time required to set up and finish the test. (8)					
The organisation of the exam/test. (9)					
The backup of students' EFL speaking performance. (10)					
Ease of the practice of this testing method. (11)					
The flexibility of this testing method. (12)					
The effectiveness of this testing method in assessing EFL speaking skills. (13)					
The feasibility of this testing method with University available resources. (14)					

Q12 Students' interest of different aspects of digital representation assessment.

	Very dissatisfied (1)	Dissatisfied (2)	Neutral (3)	Satisfied (4)	Very satisfied (5)
Management of the exam/test. (1)					
Marking of students' speaking performance. (2)					

The reliability of the test results. (3)

The validity of the assessment. (4)

The economical features of applying this testing method. (5)

The application of new technology in the exam/test. (6)

The pedagogical effects (The testing method may support and enhance EFL speaking teaching and learning). (7)

Time required to set up and finish the test. (8)

The organisation of the exam/test. (9)

The backup of students' EFL speaking performance. (10)

Ease of the practice of this testing method. (11)

The flexibility of this testing method. (12)

The effectiveness of this testing method in assessing EFL speaking skills. (13)

The feasibility of this testing method with University available resources. (14)

Q13 Two things that I like best about digital representations of EFL speaking for assessment.

_____

Q14 Two things that I do not like about digital representations of EFL speaking for assessment.

_____

Q15 Were there any technical problems with doing the activities?

Yes. (1) _____No. (2) _____

Q16 Were there other problems with the activities?

Yes. (1) _____No. (2) _____

Q17 Have you got any suggestions for improving the use of digital representations of EFL speaking for assessment?

Yes. (1) _____No. (2) _____

Q18 There will be opportunities for you to discuss with the Researcher about this new testing method. Would you like to attend an interview with the Researcher?

Yes. Your email or your phone number. (1) _____

No. (2) _____

I will contact you later. (3) _____

Appendix P: Cronbach's alpha reliability coefficient range

Value	Alpha reliability
> .9	Excellent
> .8	Good
> .7	Acceptable
> .6	Questionable
> .5	Poor
< .5	Unacceptable

(Adapted from George (2011)*)*

Appendix Q: Teacher Invitation Letter

Edith Cowan University

AUSTRALIA ECU EDITH COWAN UNIVERSITY

Invitation to participate in the Research Project:

DIGITAL REPRESENTATIONS FOR ASSESSMENT OF SPOKEN EFL AT UNIVERSITY LEVEL: A VIETNAMESE CASE STUDY

Dear FPT Teacher,

My name is Thi Bich Hiep Vu, and I am writing to you as a student of the School of Education at Edith Cowan University, Western Australia. I would like to invite you to participate in a research project I am undertaking as part of a Doctor of Philosophy in Education degree. The purpose of my research is to investigate the reliability and the feasibility of digital representations of English speaking assessment in Vietnam. The research will address the problems of low reliability of English speaking tests and potentially contribute to the improvement of oral proficiency assessment of English as a foreign language in Vietnam.

I am seeking your consent to participate in the research as invigilators and/or assessors in two phases of the research. As an invigilator, you will be asked to invigilate the practice English speaking test and do the marking of students' speaking performance in the current way – the way that you usually mark students' speaking performance at your university now. You will be observed during the test time. The invigilating will take one and a half hour. As an assessor, you will be asked to do the marking of students' digital representations of speaking performance. Students' digital representations and the marking instructions will be shared with you via email. The assessing activity will take you 30 minutes to one hour. You can choose to be an invigilator or an assessor or both. The research has no significant potential risks. Your participation in the research may take you a little time to attend the English speaking test and finish the survey and the interview. However, you will gain experience with the new speaking testing technique and have opportunity to express your opinions about different testing techniques.

After submitting students' results to the Researcher, you will complete a survey questionnaire. We anticipate the survey will take approximately 10-15 minutes. Then you will be invited to take part in a friendly interview with the Researcher. The interview will last 15-30 minutes.

You will also be asked to send my request to your students to invite them to participate in the practice English speaking test. The request will contain an information letter and a consent letter.

The information you and your students provide will be confidential and de-identified. The collected data will be used in my PhD studies, thesis and publications, and stored securely on ECU premises for ten years after the research has concluded and will then be permanently deleted.

Participation in this research is voluntary and you are free to withdraw before the test time in Phase Two if you participate as an invigilator or both, and before getting emails with students' videos in Phase Three if you participate as an assessor, and there is no penalty for doing so. If you would like to take part in the research, please sign the Consent letter and hand it to the Researcher. Your participation will ensure the success of the research.

If you have any questions, please do not hesitate to contact me**:**

Thi Bich Hiep VU

PhD candidate, School of Education

Edith Cowan University

2 Bradford St, Mount Lawley WA 6050

Tel: [redacted] or [redacted]

Email: [redacted]

You can also contact my supervisor:

Dr. Jeremy Pagram

Senior Lecturer for the School of Education

Associate Director for the Centre for Schooling and Learning Technologies

Edith Cowan University

2 Bradford St, Mount Lawley WA 6050

Tel: +61 (8) 9370 6331

Email: j.pagram@ecu.edu.au

Best regards,

Thi Bich Hiep VU

The research has been approved by the Edith Cowan University Human Research Ethics Committee. If you wish to have more information about the conduct of the research, please contact the Research Ethics Office on + 61 (8) 6304 2170 or by email research.ethics@ecu.edu.au.

Appendix R: Student Invitation Letter

Edith Cowan University

AUSTRALIA ECU EDITH COWAN UNIVERSITY

Invitation to participate in the Research Project:

DIGITAL REPRESENTATIONS FOR ASSESSMENT OF SPOKEN EFL AT UNIVERSITY LEVEL: A VIETNAMESE CASE STUDY

Dear FPT Student,

My name is Thi Bich Hiep Vu, and I am writing to you as a student of the School of Education at Edith Cowan University, Western Australia. I would like to invite you to participate in a research project I am undertaking as part of a Doctor of Philosophy in Education degree. The purpose of my research is to investigate the reliability and the feasibility of digital representations of English speaking assessment in Vietnam. The research will address the problems of low reliability of English speaking tests and potentially contribute to the improvement of oral proficiency assessment of English as a foreign language in Vietnam.

I am seeking your consent to participate in research by taking part in the practice English speaking test which is similar to the normal English test you take as part of your studies. Your participation in the research may take you a little time to attend the English speaking test and finish the survey and the interview. This test will be useful practice for you. You will get teachers' feedback and assessment results on your English speaking skills. Your marks, which you get from the practice test, will not be recorded in your school report.

During the practice test, you will be observed and videoed. The testing activity will take you 8- 10 minutes.

After the practice test, you will be asked to complete a paper survey questionnaire. We anticipate the survey will take approximately 10-15 minutes.

After teachers finish marking, you will receive your testing results and the videos of your English speaking performance. You will be invited to take part in a friendly interview. The interview will take you about 10-15 minutes.

The information you provide will be confidential and de-identified; this means that your name will not be attached to the information. The collected data will be used in my PhD studies, thesis and publications, and stored securely on ECU premises for ten years after the research has concluded and will then be permanently deleted.

Participation in this research is voluntary and you are free to withdraw before the test time, and there is no penalty for doing so. If you would like to take part in the research, please sign the Consent letter and hand it to the Researcher. Your participation will ensure the success of the research.

If you have any questions, please do not hesitate to contact me**:**

Thi Bich Hiep VU, PhD candidate, School of Education, Edith Cowan University

2 Bradford St, Mount Lawley WA 6050. Tel: [redacted] or [redacted]

Email: [redacted]

You can also contact my supervisor:

Dr. Jeremy Pagram, Senior Lecturer for the School of Education

Associate Director for the Centre for Schooling and Learning Technologies

Edith Cowan University

2 Bradford St, Mount Lawley WA 6050. Tel: +61 (8) 9370 6331

Email: j.pagram@ecu.edu.au

Best regards,

Thi Bich Hiep VU

The research has been approved by the Edith Cowan University Human Research Ethics Committee. If you wish to have more information about the conduct of the research, please contact the Research Ethics Office on + 61 (8) 6304 2170 or by email. research.ethics@ecu.edu.au.

Appendix S: Comparison of textbooks to International Standards and Tests

	International Standards	TOEFL (Paper/iBT)	IELTS	CEF
Summit 1	High-Intermediate	525-575/ 70-90	5.0	B2/Level 3
Top Notch 3	Intermediate	475-525/ 52-70	4.0	B1/Level 2
Top Notch 2	Pre-Intermediate	425-475/ 38-52	3.0	A2/Level 1

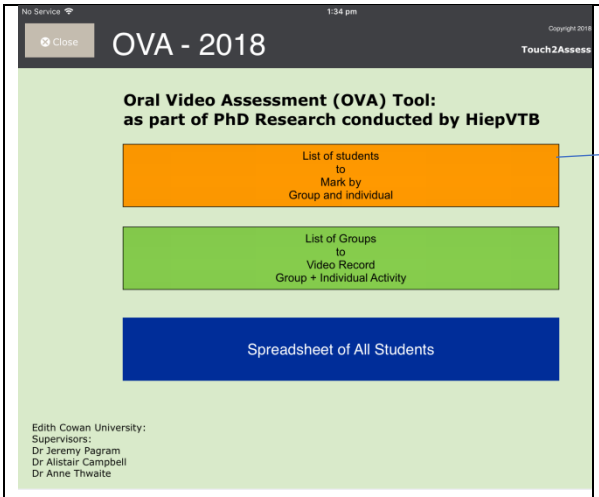
Source:

http://www.pearsonlongman.com/summit/downloads/correlations/TN_Summit_corrs_intltests.pdf

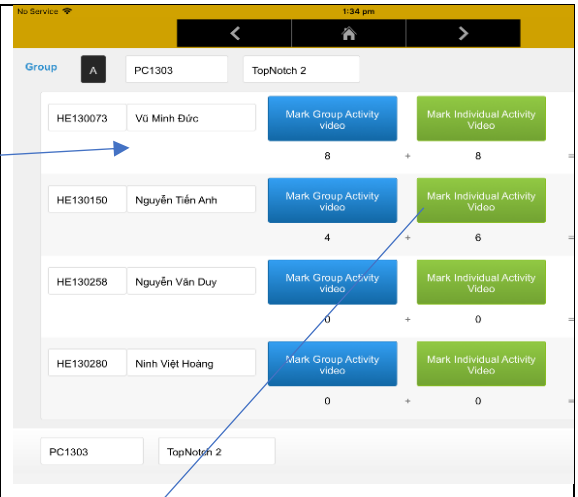
Appendix T: Marker guideline

MARKER GUIDELINE

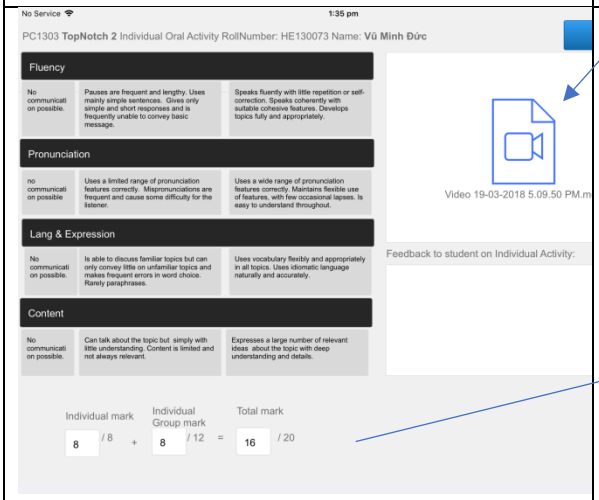
iPad password: 6876 Software Username: OVA Software password: O



The Assessment Tool Interface

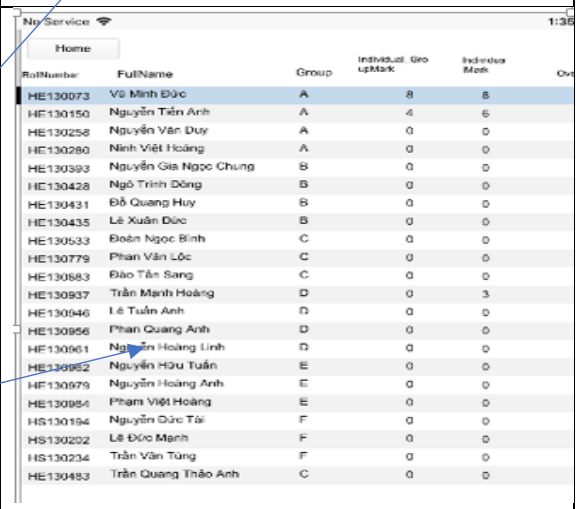


Home, Backward, Forward buttons help you move around.

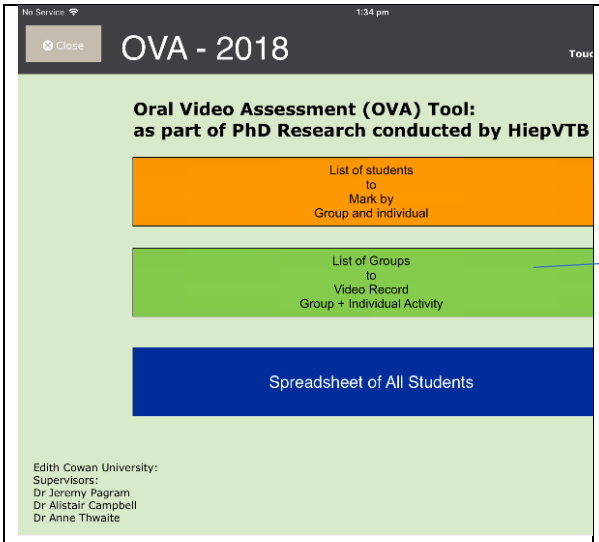


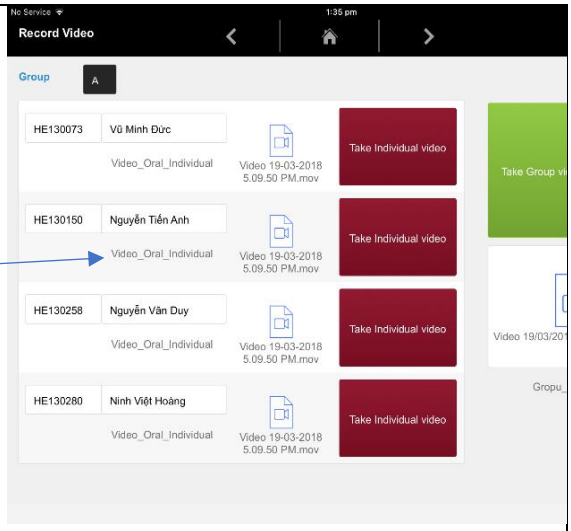
Click  , choose Play Video to watch students' videos.

Click on a particular key, and students' marks will be added up and recorded automatically.



The Spreadsheet can be printed out or sent to teachers' email.





The Assessment Tool Interface	This is how to video students' performance with maximum time pre-set.

Oral Video Assessment – 2018

Guideline prepared by Thi Bich Hiep VU – PhD candidate, Edith Cowan University.

Appendix U: The Public version IELTS Speaking Band Descriptor

IELTS

SPEAKING: Band Descriptors (public version)

Band	Fluency and coherence	Lexical resource	Grammatical range and accuracy	Pronunciation
9	• speaks fluently with only rare repetition or self-correction; • any hesitation is content-related rather than to find words or grammar • speaks coherently with fully appropriate cohesive features • develops topics fully and appropriately	• uses vocabulary with full flexibility and precision in all topics • uses idiomatic language naturally and accurately	• uses a full range of structures naturally and appropriately • produces consistently accurate structures apart from 'slips' characteristic of native speaker speech	• uses a full range of pronunciation features with precision and subtlety • sustains flexible use of features throughout • is effortless to understand
8	• speaks fluently with only occasional repetition or self-correction; hesitation is usually content-related and only rarely to search for language • develops topics coherently and appropriately	• uses a wide vocabulary resource readily and flexibly to convey precise meaning • uses less common and idiomatic vocabulary skilfully, with occasional inaccuracies • uses paraphrase effectively as required	• uses a wide range of structures flexibly • produces a majority of error-free sentences with only very occasional inappropriacies or basic/non-systematic errors	• uses a wide range of pronunciation features • sustains flexible use of features, with only occasional lapses • is easy to understand throughout; L1 accent has minimal effect on intelligibility
7	• speaks at length without noticeable effort or loss of coherence • may demonstrate language-related hesitation at times, or some repetition and/or self-correction • uses a range of connectives and discourse markers with some flexibility	• uses vocabulary resource flexibly to discuss a variety of topics • uses some less common and idiomatic vocabulary and shows some awareness of style and collocation, with some inappropriate choices • uses paraphrase effectively	• uses a range of complex structures with some flexibility • frequently produces error-free sentences, though some grammatical mistakes persist	• shows all the positive features of Band 6 and some, but not all, of the positive features of Band 8
6	• is willing to speak at length, though may lose coherence at times due to occasional repetition, self-correction or hesitation • uses a range of connectives and discourse markers but not always appropriately	• has a wide enough vocabulary to discuss topics at length and make meaning clear in spite of inappropriacies • generally paraphrases successfully	• uses a mix of simple and complex structures, but with limited flexibility • may make frequent mistakes with complex structures though these rarely cause comprehension problems	• uses a range of pronunciation features with mixed control • shows some effective use of features but this is not sustained • can generally be understood throughout, though mispronunciation of individual words or sounds reduces clarity at times
5	• usually maintains flow of speech but uses repetition, self correction and/or slow speech to keep going • may over-use certain connectives and discourse markers • produces simple speech fluently, but more complex communication causes fluency problems	• manages to talk about familiar and unfamiliar topics but uses vocabulary with limited flexibility • attempts to use paraphrase but with mixed success	• produces basic sentence forms with reasonable accuracy • uses a limited range of more complex structures, but these usually contain errors and may cause some comprehension problems	• shows all the positive features of Band 4 and some, but not all, of the positive features of Band 6
4	• cannot respond without noticeable pauses and may speak slowly, with frequent repetition and self-correction • links basic sentences but with repetitious use of simple connectives and some breakdowns in coherence	• is able to talk about familiar topics but can only convey basic meaning on unfamiliar topics and makes frequent errors in word choice • rarely attempts paraphrase	• produces basic sentence forms and some correct simple sentences but subordinate structures are rare • errors are frequent and may lead to misunderstanding	• uses a limited range of pronunciation features • attempts to control features but lapses are frequent • mispronunciations are frequent and cause some difficulty for the listener
3	• speaks with long pauses • has limited ability to link simple sentences • gives only simple responses and is frequently unable to convey basic message	• uses simple vocabulary to convey personal information • has insufficient vocabulary for less familiar topics	• attempts basic sentence forms but with limited success, or relies on apparently memorised utterances • makes numerous errors except in memorised expressions	• shows some of the features of Band 2 and some, but not all, of the positive features of Band 4
2	• pauses lengthily before most words • little communication possible	• only produces isolated words or memorised utterances	• cannot produce basic sentence forms	• Speech is often unintelligible
1	• no communication possible • no rateable language			
0	• does not attend			

Source: https://www.ielts.org/-/media/pdfs/speaking-band-descriptors.ashx?la=en